

Essays in the Economics of Education

A DISSERTATION PRESENTED BY

JASON COOK

TO

THE ECONOMICS DEPARTMENT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

ECONOMICS

CORNELL UNIVERSITY

ITHACA, NEW YORK

May 2017

Copyright® 2017 by Jason Cook

All rights reserved

ESSAYS IN THE ECONOMICS OF EDUCATION

Jason Cook, Ph.D.

Cornell University 2017

ABSTRACT

This dissertation is comprised of three essays, each studying a different aspect related to topics in the economics of education. These pieces each study how education production is impacted by different economic forces. In Chapter 1, I study the effect of racial segregation on academic achievement, college preparation, and postsecondary attainment in a large, urban school district. To achieve racial balance in its oversubscribed magnet schools, this district conducted separate admissions lotteries for black and non-black students. Because the student body was predominantly black, administrators set aside disproportionately more seats for the non-black lottery. In 2003, the federal Office of Civil Rights forced this district to instead use a race-blind lottery procedure that dramatically increased racial segregation for incoming magnet school cohorts. In an instrumental variables framework that exploits both randomized lottery offers and this unanticipated shock to racial makeup, I test whether student racial composition is a meaningful input in the education production function. As a baseline, I use admissions lotteries to estimate the effect of enrolling in a magnet middle school on student outcomes. In general, enrollment returns are comparable between magnet and traditional schools, but I estimate heterogeneous magnet school effects across student subgroups. Education production is sensitive to school racial composition in that segregation has a deleterious impact on student outcomes. I find that increasing the share of black peers in a cohort decreases student achievement in math, science, and writing for black students with losses primarily driven by high-aptitude black students. Further, racial segregation erodes high school graduation rates and also decreases college attendance by reducing enrollment at 2-year institutions among female black students. These findings suggest that policies aimed at achieving racial balance in schools will likely increase aggregate educational achievement.

Chapter 2 examines the impact of competition due to charter school entry on the level and composition of expenditures within traditional public school districts (TPSDs). I leverage policy changes affecting the location and timing of charter entry to account for endogenous charter competition. TPSDs respond to competition by allocating resources away from instructional and other expendi-

tures towards new capital construction. Using teacher contracts, I show the declines in instructional spending are partially due to decreases in collectively bargained salaries. Competition depresses appraised housing valuations, in turn causing TPSDs to lose property tax revenues resulting in a decline in overall spending.

In Chapter 3, which is joint work with Richard Mansfield, we use administrative panel data to decompose worker performance into components relating to general talent, task-specific talent, general experience, and task-specific experience. We consider the context of high school teachers, in which tasks consist of teaching particular subjects in particular tracks. Using the timing of changes in the subjects and difficulty levels to which teachers are assigned to provide identifying variation, we show that a substantial part of the productivity gains to teacher experience are actually subject-specific. Similarly, while three-quarters of the variance in the permanent component of productivity among teachers is portable across subjects and levels, there exist non-trivial subject-specific and level-specific components. Counterfactual simulations suggest that maximizing the test-score contribution of task-specific experience and task-specific talent can increase student performance by as much as .04 test score standard deviations relative to random assignment of teachers to classrooms.

BIOGRAPHICAL SKETCH

Jason received a Bachelor's degree in economics and mathematics from Brigham Young University and a Master's degree in economics from Cornell University.

ACKNOWLEDGMENTS

I am indebted to my dissertation advisors Francine Blau, Michael Lovenheim, and Jordan Mat-sudaira. My time researching for Fran provided my foundation for doing careful and thoughtful research. Her high expectations and guidance were instrumental in shaping my research interests as well as preparing me for life in academia. Mike always made time to meet to discuss ideas and brainstorm solutions to problems I encountered. His thoughtful and detailed feedback and revisions not only had a large impact on this current work, but the skills I have picked up from my time with him has shaped my future research agenda. Jordan also provided insightful feedback and was always able to identify the crucial issues that needed addressing in my work. His applied econometrics class was so thoughtful and well executed that I find myself drawing from lessons I learned on a daily basis. My entire committee provided tireless feedback on the many drafts of this work and were instrumental in shaping the final product. They all went out of their way to prepare me for life in academia and I am extremely grateful for their time.

Financial support was graciously provided by the National Academy of Education/Spencer Dissertation Fellowship Program. I would like to thank Andrew Johnston, Rich Patterson, Jeffrey Swigert, Corbin Miller, and Todd Jones for helpful academic discussions and great friendship.

Finally, I thank my wife Eliza, whose support made this dissertation possible. I cannot count how many late nights we spent on campus preparing for qualifying exams and working against the never ending series of deadlines that accompany a PhD. She not only has made finishing the dissertation possible, but made it one of the happiest times of my life. I dedicate this dissertation to my family, including our new recent addition, Aletta Lark.

Contents

1 Segregation, Student Achievement, and Postsecondary Attainment: Evidence from the Introduction of Race-Blind Magnet School Lotteries	1
1.1 Introduction	1
1.2 Prior Literature	6
1.3 Institutional Details	10
1.3.1 Magnet Schools in this Large Urban School District	10
1.3.2 No Child Left Behind	12
1.4 Data	13
1.5 Estimating the Baseline Returns to Magnet School Attendance	15
1.5.1 Magnet School Lottery	15
1.6 Estimating Peer Racial Composition Effects	17
1.6.1 Validating the Instrumental Variables Strategy	21
1.6.2 Differential Attrition	23
1.6.3 Accounting for No Child Left Behind and Changes in the Composition of Applicants	24
1.7 Magnet Enrollment Effects	26
1.7.1 Effects of Magnet Enrollment on Teacher and Peer Characteristics	26
1.7.2 Effects of Magnet Enrollment on Student Outcomes	28
1.8 The Effect of Segregation on Student Outcomes	30
1.9 Conclusion	33
1.10 Figures and Tables	35

2	The Effect of Charter Competition on Unionized District Revenues and Resource Allocation	49
2.1	Institutional Details	52
2.2	Data	53
2.2.1	Data Description	53
2.2.2	Measuring Charter Competition	55
2.3	Methodology	57
2.3.1	Baseline Estimates: Difference-in-Difference Framework	57
2.3.2	Preferred Estimates: Instrumental Variables Framework	58
2.3.3	Adjusting Methodology for Contract Outcomes	64
2.3.4	Validating the Instrumental Variables Strategy	67
2.4	Student and Teacher Mobility Responses	69
2.5	District Revenues	71
2.6	Collectively Bargained Teacher Contracts	76
2.7	Resource Allocation	78
2.8	Discussion	79
2.9	Conclusion	81
3	Task-Specific Experience and Task-Specific Talent: Decomposing the Productivity of High School Teachers	94
3.1	Introduction	95
3.2	Model Specification	101
3.3	Identification	104
3.3.1	Identifying the Returns to General and Task-Specific Experience	104
3.3.2	Identification of the General and Context-Specific Components of Fixed Teaching Skill	107
3.4	Data	113
3.4.1	Overview	113
3.4.2	Task-Specific Output and Sample Restrictions	113
3.4.3	Generating the Experience Profile	116

3.4.4	The Frequency of Teacher Assignment Rotations	117
3.4.5	Estimation and Calculation of Standard Errors	119
3.5	Results	119
3.5.1	Variation in the General and Context-Specific Components of Time-Invariant Teacher Productivity	119
3.5.2	General and Context-Specific Experience Profiles	121
3.6	Tests of Identifying Assumptions and Robustness Checks	124
3.6.1	Testing for Dynamic Classroom Assignment Responses to Unobserved Shocks	124
3.6.2	Testing for Dynamic Student Sorting	126
3.6.3	Evaluating Forecast Bias in Estimates of Context-Specific Teacher Talent . .	128
3.6.4	Further Robustness Checks	130
3.7	Projecting the Achievement Gains from Efficient Use of Context-Specific Teacher Experience and Talent	135
3.7.1	Methodology	135
3.7.2	Results from Counterfactual Simulations	139
3.8	Conclusions	144
3.9	Tables and Figures	145
I	Appendices	162
4	Segregation Appendix	164
A	Differential Trends Test: Main Outcomes	164
B	Disaggregated Magnet Effects	167
B.1	Effects of Magnet Enrollment on Academic Achievement	167
B.2	Effects of Magnet Enrollment on ACT Testing	172
B.3	Effects of Magnet Enrollment on Postsecondary Attainment	179
C	First-Stage Estimates, Observation Counts, and Outcome Averages	182
5	Charter Competition Appendix	191
A	Data Appendix	191

A.1	ODE Restricted-Access Staff Data	191
A.2	SERB Collectively Bargained Contract Data	192
A.3	Other Sample Restrictions	192
B	Full Salary Distribution Imputation	195
C	Alternative Measures of Charter Competition	195
D	Ohio School Rating Designation System	198
D.1	State Indicators	198
D.2	Performance Index Scores	198
D.3	School and District Rating Calculations	198
D.4	Value Added	199
D.5	Ratings Determined by Number of AYP Indicators met	199
E	Robustness: No Child Left Behind & Great Recession	203
F	Mechanical Bias for Models with Partially Fixed Dependent Variables	208
F.1	Monte Carlo Simulations	210
G	Unobservable Trend IV Check: All Outcomes	213
H	Regression Overidentification Tests and Outcome Means	220
6	Teacher Experience Appendix	222
A	Identification of Experience Profiles	222
B	Identification of Permanent Teaching Skill	224
C	Recovering the Latent Variance Decomposition	224
D	Testing the Additive Separability of Context-Specific Experience Profiles	226
D.1	Smoothing the Nonparametric Experience Contribution Function	226
D.2	Marginal Effects Example	226
E	Methodology for Measuring Forecast Bias	227
F	Formulation of the Counterfactual Simulation	229

Chapter 1

Segregation, Student Achievement, and Postsecondary Attainment: Evidence from the Introduction of Race-Blind Magnet School Lotteries

1.1 Introduction

The landmark ruling of *Brown v. Board of Education of Topeka* ended *de jure* school segregation and spurred integration efforts across the United States education system. The assumption underlying this significant ruling is that peer racial composition is a meaningful parameter in the education production function. Specifically, the ruling assumes that racial isolation negatively affects student outcomes, particularly for minorities. However, the effects of peer racial composition on academic outcomes are still not known. Despite early integration efforts, schools nationwide are growing increasingly *de facto* segregated (Lutz, 2011; Reardon et al., 2012; GAO, 2016; Clotfelter et al., 2006, 2008). In the 2013-2014 school year, over 6.5 million students attended schools in which over 90 percent of their peers were black or Hispanic.¹ Moreover, the proportion of these high-minority-share schools have tripled over the last two decades nationwide (Orfield et al., 2016). With *de*

¹Author's calculations using Common Core Data from the National Center for Education Statistics.

facto school segregation on the rise, the causal link between peer racial composition and student achievement has important implications for policy.

I directly test whether peer racial composition is a meaningful input in the education production function by studying the end of race-conscious admissions lotteries in a large urban school district (LUSD). The change in the lottery regime caused magnet middle schools that were nearly racially balanced to instead enroll a high share of minority students. Thus, more specifically, I explore how a large, exogenous increase in racial segregation impacts education production in the context of magnet middle schools.

Magnet schools provide an ideal setting to explore the impact of racial segregation on academic outcomes. Magnets were established as a voluntary alternative to compulsory desegregation efforts such as busing. While being publicly funded and operated, magnets differ from traditional public schools in that they are permitted to offer specialized programs and services. They also differ in that they lack specified catchment boundaries allowing them to attract enrollment district-wide, hence the term “magnet.” In theory, districts attempting to discourage racial segregation would establish magnet schools touting specialized programs within high-minority-share neighborhoods to encourage non-resident white families to enroll their children. Thus, magnets promote racial balance in what would otherwise be high-minority-share schools.² Despite racial balance being a founding principle motivating the creation of magnet schools, we have no understanding about the extent to which racial composition drives magnet school achievement gains.

I begin by establishing the baseline effect of enrolling in a magnet school on achievement and postsecondary attainment using two decades of LUSD admissions lotteries. To my knowledge, this is the longest panel of lotteries used in any admissions lottery study to date.

I then isolate the impact of segregation on the effectiveness of magnet enrollment using an instrumental variables design based on a change in the district’s lottery system. Prior to 2003, the LUSD ran magnet school lotteries separately for black and non-black students. This provided district administrators full control over the racial composition of each magnet school’s entering class, allowing them to artificially improve racial balance in their admissions by providing disproportion-

²However in practice, [Rossell \(2003\)](#) finds that adding voluntary magnet programs to a district’s desegregation plan has little impact on exposure to other races.

ately more offers to non-black students. In 2003, the federal Office of Civil Rights required the district to consolidate their race-specific lotteries to a system with a combined, race-blind lottery. Under this regime, the racial composition of the entering class simply mirrored the racial makeup of the lottery pool. Importantly, the lottery consolidation only affected the admissions process and left other school policies and staffing unaffected.³ Thus, the shift in the racial composition of a school’s incoming class induced by the lottery consolidation was a function of how disproportionately non-black the pre-consolidation lottery winners were compared to the racial composition of the school’s entire lottery pool.

Following this intuition, my identification strategy isolates the exogenous variation in magnet school racial composition that is induced by the district’s lottery consolidation. I instrument for the racial composition in a student’s enrolled school using a measure of how “disproportionately non-black” the school’s lottery winners were prior to the consolidation interacted both with indicators for whether the student won a magnet lottery and whether the lottery occurred after the consolidation. The main threat to this strategy is if unobserved determinants of the effect of magnet offers on student outcomes are trending differentially for magnets with more “disproportionately non-black” lottery offers in 2002. I test this assumption in an event study framework and find no evidence for the existence of such trends.

The other main threat to the validity of this strategy is if changes to the composition of the lottery pool after 2002 are correlated with “disproportionately white offers” for magnet schools. I test for changes in the composition of the lottery pool and find that magnet schools with more “disproportionately non-black offers” have a higher proportion of black and male applicants following the termination of race-based admissions. However, I find no evidence for such compositional changes in baseline achievement both with and without conditioning on student race and gender. Thus, by exploring heterogeneous effects of segregation by race and gender subgroups, I remove these compositional effects and can isolate the impact of segregation on student outcomes.

My baseline estimates reveal that the returns to magnet middle school enrollment are generally

³Staff reshuffling could result from changes in student demographics (Jackson, 2009), however, none of these changes were structurally linked with the lottery consolidation and can be considered part of the segregation treatment.

statistically indistinguishable from traditional school enrollment. However, magnet middle schools boost science achievement for non-black students as well as female students and increase ACT test taking among non-black as well as low-achieving students. The localized nature of the returns to enrolling in a magnet school relative to a traditional public school highlights the similarities between both institutions. Thus, I argue that any effects of segregation I estimate within magnet schools may generalize to traditional schools more broadly.

The end of race-conscious admissions lotteries led to an immediate 7 percentage point increase in segregation among magnet middle schools as measured by the exposure index (Massey and Denton, 1988).⁴ This is slightly larger than the immediate change in exposure index resulting from the end of forced busing in Charlotte-Mecklenburg (Billings et al., 2014) or roughly half the effect of court-ordered desegregation in the 1960s and 70s (Rossell and Armor, 1996; Guryan, 2004). Racial segregation in magnet schools has deleterious effects on student outcomes. A 10 percentage point increase in the share of black peers at a student’s school, which represents an increase in racial segregation, decreases student achievement by roughly 0.12 standard deviations. These effects are slightly larger than other estimates in the literature (Hoxby, 2000; Hanushek et al., 2009; Billings et al., 2014). Similar to Hanushek and Rivkin (2009), I find that the losses from segregation are concentrated among high-achieving black students. Segregation in magnet middle schools also decreases the probability of graduating from high school and later enrolling in a two-year college for black female students. I conclude that racial balance is an important input into the education production function in the magnet schools I study.

This paper makes contributions to two different literatures. First, I add to the literature studying the returns to magnet schools by providing the first lottery evidence of magnet middle school attendance on postsecondary attainment. Second, I contribute to the literature studying the effect of school segregation and peer racial composition on student academic outcomes by leveraging a natural experiment that is better suited to isolate the contribution of peer racial composition on education production. Previous studies have assessed changes to peer racial composition driven by naturally-occurring variation in cohort- or classroom-specific racial composition (Hoxby, 2000; Vigdor and Nechyba, 2007; Hanushek et al., 2009; Hanushek and Rivkin, 2009). However, by

⁴The exposure index measures the probability that a randomly chosen peer of a minority student is also a minority.

construction, these strategies exploit very small differences in peer composition, which plausibly affect student achievement differently than policy reforms that generate large changes to peer composition.⁵ Other studies assess policies that induce large shifts to peer composition including the introduction of court-ordered desegregation (Guryan, 2004; Johnson, 2015) or its termination (Billings et al., 2014; Lutz, 2011; Gamoran and An, 2016), inner-city busing (Angrist and Lang, 2004), a change in attendance zone boundaries (Vigdor and Nechyba, 2007; Billings et al., 2014), or mandated school reassignment (Hoxby, 2006). However, these studies do not occur in settings where students are explicitly randomly assigned to schools. Because my empirical strategy leverages explicit randomization into schools both before and after a large shift in school segregation, I am able to provide a cleaner estimate than previous work of how school racial composition enters into the education production function.⁶

My findings are highly relevant for current education policy along a number of dimensions given the resurgence of *de facto* segregation over the past few decades in the United States. Because racial balance appears to be an important factor in the education production of magnet schools, districts may have additional justification to implement policies that promote racial balance.⁷ In fact, President Obama’s “Stronger Together” initiative currently proposes to double the amount of federal funding up to \$120 million to improve voluntary integration efforts across the United States, of which magnet programs are a part.

I argue that the negative effect of segregation is likely not specific to the magnet school setting. Magnet schools in this district are statistically comparable to traditional schools with regard to

⁵Vigdor and Nechyba (2007) estimate racial composition peer effects using classroom-specific variation, but conversely find no evidence for racial peer effects when limiting the analysis to year-to-year variation induced by changes to school assignment policies.

⁶Two recent studies assess the effect of peer racial composition using regression discontinuity methods based on entrance exams scores to attend highly selective exam schools in Boston and NYC (Abulkadiroglu et al., 2014; Dobbie and Fryer, 2014). Students narrowly gaining admission experience a drastically less racially diverse peer group than narrowly failing students. However, unlike in my setting, these works are unable to disentangle the effect of peer racial composition from the effect of attending the exam school (e.g., having access to better teachers or academic resources).

⁷Enforcing racial diversity in schools is complicated due to recent court cases, such as *Parents Involved in Community Schools v. Seattle School District No. 1* in 2007, which prevents districts from utilizing race in admissions decisions.

school inputs, peer composition, and general returns. Moreover, my identification strategy is able to isolate the effect of segregation from any magnet-school-specific inputs. Thus, the deleterious effect of segregation on academic outcomes plausibly generalizes to other school settings.

1.2 Prior Literature

Why might segregation impact student outcomes? One potential explanation involves the direct influence of a student’s peer group. In Florida, [Carrell and Hoekstra \(2010\)](#) find that black and low-income students are more likely to experience domestic violence at home and are more likely to be disruptive in the classroom as a result. Further, [Carrell and Hoekstra \(2010\)](#) find that exposure to these disruptive peers decreases reading and math test scores for students in this district.⁸ If these findings hold in this LUSD, then increasing the share of black students in schools may increase the probability of exposure to disruptive peers.

In a similar vein, because black students in this district have lower average test scores compared to non-black students, increasing segregation will decrease the average baseline achievement of a student’s peer group. There is a large, mixed literature testing for the presence of peer achievement effects in schools (see [Sacerdote, 2011](#), for a detailed review). Also, if teachers adjust how they teach to the aptitude of the average student, then decreasing peer baseline achievement could affect the teacher’s contribution to student outcomes ([Duflo et al., 2011](#)).

It is impossible to separately isolate the contribution of these different potential mechanisms within my setting. Instead, this study provides estimates for the combined impact of these channels on student outcomes. However, my reduced-form estimates of segregation effects are inherently interesting to researchers and policy makers faced with evaluating the impact of a policy that will influence school racial composition.

This study contributes to two different literatures. The first is a growing, but mixed literature estimating the short- and medium-run academic returns to attending magnet schools using lotteries. Several studies estimate limited-to-no academic returns to magnet attendance ([Cullen et al., 2006](#);

⁸However, [Hoxby \(2006\)](#) find little support for “bad apple” models of peer effects ([Lazear, 2001](#)) in North Carolina.

Cullen and Jacob, 2007; Engberg et al., 2014; Abulkadiroglu et al., 2014; Dobbie and Fryer, 2014; Dee and Lan, 2015), while other studies estimate academic returns roughly half the magnitude of lottery-based charter school estimates (Hastings et al., 2012; Bifulco et al., 2009; Crain et al., 1992).⁹ However, we know far less about long-run returns to magnet enrollment particularly in the United States.¹⁰ Several studies explore the short-, medium-, and long-run returns to open enrollment systems, of which magnet schools are a part (Hastings et al., 2006, 2009; Deming, 2011; Deming et al., 2014). However, these studies do not separately report estimates for magnet enrollment. Hoxby (2003a) asserts that magnet schools should not be considered as part of the school choice movement. She argues that magnet schools predate the school choice movement and do not provide the same financial incentive structure as traditional school choice programs, which further motivates separately exploring the contribution of magnets. I contribute to this literature by providing the first estimates of the effect of magnet middle school enrollment on postsecondary attainment using administrative lottery data.

Second, this work contributes to the large literature assessing the impact of segregation and racial peer effects on student outcomes.¹¹ While there are studies that estimate little-to-no effect of student racial composition on achievement (Gamoran and An, 2016; Hoxby, 2006; Abulkadiroglu et al., 2014; Dobbie and Fryer, 2014), many others find that increasing the share of minority peers negatively impacts student achievement and behavioral outcomes particularly among minority subgroups and females (Vigdor and Nechyba, 2007; Angrist and Lang, 2004; Billings et al., 2014; Lutz, 2011; Hoxby, 2000; Hanushek et al., 2009; Guryan, 2004; Hanushek and Rivkin, 2009).¹²

⁹Cullen et al. (2006) and Engberg et al. (2014) estimate limited academic returns to magnet enrollment, but do find behavioral effects. Crain and Thaler (1999) find positive effects for some types of magnets and null or negative effects for other types.

¹⁰Crain and Thaler (1999) provide qualitative evidence about postsecondary attainment and compare in-depth survey responses between lottery winners and losers for 110 students. Park et al. (2015) find that magnet attendance increases the probability of attending college in rural China.

¹¹See Gamoran and An (2016) for a full review of the literature estimating the effect of segregation on student achievement and Vigdor and Ludwig (2008) for a review on the literature relating neighborhood and school segregation to the black-white test score gap.

¹²Hoxby (2006) find evidence that peer race and ethnicity have only slight effects once conditioning on peer achievement. Vigdor and Nechyba (2007) find that school-wide racial composition does not significantly predict achievement, however, they estimate that non-black students in classrooms with a disproportionately share of black

To overcome selection biases, several studies rely on quasi-random variation in racial composition generated from naturally occurring cohort- or classroom-specific variation (Hoxby, 2000; Vigdor and Nechyba, 2007; Hanushek et al., 2009; Hanushek and Rivkin, 2009). However, the quasi-randomization that allows these studies to address selection concerns also reduces variation in racial composition. Thus, these studies identify effects from small fluctuations in racial composition, which potentially impact student outcomes differently than a large, policy-induced shift in racial composition. Other studies utilize policies that induce large shifts in racial composition such as the introduction of court-ordered desegregation (Guryan, 2004; Johnson, 2015) or its termination (Billings et al., 2014; Lutz, 2011; Gamoran and An, 2016), inner-city busing (Angrist and Lang, 2004), the change in attendance zone boundaries (Vigdor and Nechyba, 2007; Billings et al., 2014), or mandated school reassignment (Hoxby, 2006). These studies lack the benefits that exploiting randomized peer composition provides for identification, though each study goes to great lengths to show that identification assumptions are met.

The work of Billings et al. (2014) is most closely related to this paper. Billings et al. (2014) study the effect of segregation induced by the end of forced busing in Charlotte-Mecklenburg. They compare students who live in the same neighborhood and school zone prior to the end of forced busing, but then live on opposite sides of newly drawn school catchment boundaries, and thus, go on to attend schools with drastically different peer racial compositions. They find that segregation decreases high school achievement for white and minority students as well as lowers graduation rates and college attendance among white students and increases crime among minority males. My study compliments this seminal work in several ways. First, schools in my setting were not allocated compensatory resources due to increased segregation, which Billings et al. (2014) show may have mitigated segregation effects for younger cohorts in their study.¹³ Second, because students in my setting are explicitly randomly assigned to schools both before and after the segregation treatment, my natural experiment is better situated to cleanly isolate racial composition effects.¹⁴ Finally,

students experience lower math achievement. Johnson (2015) finds that school desegregation positively impacts adult outcomes and that school funding is the likely mechanism as opposed to any direct effects of changing the racial composition of a student’s peers.

¹³Reber (2010) also shows that desegregation effects are attributed to increased resources rather than peer racial composition.

¹⁴For example, Billings et al. (2014) find that white students who are assigned to school zones with a higher

because this article and the work of [Billings et al. \(2014\)](#) are studying different policies, both studies generate policy implications better suited to their respective settings. The findings of [Billings et al. \(2014\)](#) are more relevant to assessing the effects of a policy that ends forced busing and changes school zone assignment, while the results in my setting are more relevant to understanding the implications of a policy that ends race-conscious lotteries.

Two recent papers have explored the effect of peer composition on student outcomes using regression discontinuity evidence. These studies compare students near the admissions cutoffs to top exam schools in Boston and New York where the composition of peers are drastically different for students who are provided or denied admission ([Abulkadiroglu et al., 2014](#); [Dobbie and Fryer, 2014](#)).¹⁵ Unlike the previous work in this literature, these studies exploit both random variation in test scores near the admissions threshold as well as markedly different peer compositions experienced by treated and untreated students. However, it is difficult to distinguish peer effects from the effect of exposure to the exam school teachers and other exam-school-specific effects. Further, students who narrowly gain and lose admission will be near the bottom and top of the baseline achievement distributions in their respective schools. Thus, these regression discontinuity studies potentially conflate any effect of a student’s class ranking with peer effects.

I contribute to the literature studying segregation and racial peer effects by providing a cleaner estimate than previous work of how school racial composition enters into the education production function. My study strikes a balance between the studies exploiting explicit random variation in peer composition and those with large, policy-driven shocks to peer composition. Because students are randomized into magnet schools, my empirical strategy benefits from the virtues of randomization, while simultaneously leveraging a large shift in peer composition due to the end of race-based admissions. Further, relative to the regression discontinuity studies, my work is able to study heterogeneous effects across baseline student aptitude and does not suffer from conflating any class ranking effects.¹⁶ Additionally, I am able to isolate the effect of peer composition from

minority share are more likely to attend a magnet program. They note that if the relative returns to magnet attendance are positive then this would place upward pressure on segregation effects for white students.

¹⁵Exam schools are highly selective magnet schools with strict admissions cutoffs.

¹⁶Because the 6 exam school cutoffs hit at different parts of the student baseline achievement distribution, the estimates in [Abulkadiroglu et al. \(2014\)](#) reflect both high- and moderate-ability students.

magnet-specific effects because the segregation effects I estimate are identified off of changes in the returns to magnet enrollment across the policy change.

1.3 Institutional Details

1.3.1 Magnet Schools in this Large Urban School District

Magnet schools are similar to traditional schools in that they are publicly funded and run. All LUSD schools use the same general curriculum, but magnet schools can differ in the instruction methods used. Magnets can also emphasize a particular focus of instruction, e.g., performing arts, bilingual education, STEM, or International Baccalaureate programs. Magnet schools also differ in that they lack specified catchment boundaries allowing them to attract enrollment district-wide, hence the term “magnet.” In addition to the district’s traditional public schools, the LUSD ran roughly 10 to 15 magnet middle schools throughout the time period of this study.¹⁷

As was the case with magnet school programs across the United States, a founding principal underlying this LUSD’s magnet program was to improve racial balance and prevent “white flight.”¹⁸ Shortly after the first LUSD magnets launched in the 1970s, the district also began the mandatory busing of subsets of students to desegregate schools through the 1990s. During this time, the magnet program coexisted with forced busing as an effort to discourage middle class families from migrating to the suburbs.

Because the demand for these magnet schools far outpaced supply, magnet seats were filled via randomized lotteries. To ensure racial balance, the district held separate school-specific lotteries for black and non-black students.¹⁹ Each year the district set a universal target for the racial composition of new enrollment that reflected the racial make-up of the district as a whole. The district then set admissions quotas for each race-specific lottery to hit the district-wide target. Black students disproportionately applied to magnet schools and so students in the non-black

¹⁷Exact magnet counts are purposefully withheld to maintain the anonymity of the district.

¹⁸See [Rossell \(2005\)](#) for a detailed history of the emergence of magnet schools in the United States.

¹⁹This was a common practice for over-subscribed magnet schools across the nation. Chicago, for example, ran separate lotteries based on both gender and race ([Cullen et al., 2006](#)).

magnet lottery had a better chance at receiving a seat offer than students in the black magnet lottery. In the 2002-03 school year, the federal Office of Civil Rights required the LUSD to instead utilize a combined, race-blind lottery system comparable to the system in Durham County, North Carolina (Clotfelter et al., 2008).²⁰

In addition to filling magnet school seats with lotteries, the LUSD allowed students to apply to transfer to other oversubscribed traditional schools via the same centralized lottery. Students applied to up to three schools and did not specify a rank ordering. Lotteries would occur at the school-grade level. Once offers were made, students had roughly one week to respond. If parents failed to respond, the seat was forfeited to the next waitlisted student. Conversely, if a student accepted a seat in a school, they were automatically withdrawn from all other waitlists. Once the district was notified that an offered seat had been declined, subsequent offers were determined by moving down a randomized waitlist.²¹

To explore how race-blind lotteries impact the racial composition within LUSD schools, Figure 1.1 presents the percentage of black students enrolled in traditional and magnet schools across the time period of the study, 1998 to 2007. From 1998 to 2002, even despite utilizing race-specific lotteries, magnet schools enrolled a higher proportion of black students than traditional schools.²² Upon the introduction of race-blind lotteries in 2003-04, district administrators lost their control over the racial balance of admissions resulting in roughly a 7 percentage point increase in the black-share within magnet schools over the next few years. This also equates to roughly a 7 percentage point increase in the exposure index (Massey and Denton, 1988), which is slightly larger than the immediate increase in the exposure index due to the end of forced busing in Charlotte-Mecklenburg

²⁰The LUSD moved to a simple race-blind lottery as opposed to a race-neutral, place-based system such as with Chicago Public Schools (Ellison and Pathak, 2016) where student need is instead determined using aggregated residential neighborhood information. Other districts approached achieving race-blind balance within schools by instead incorporating information about student socioeconomic status and achievement as in Wake County, North Carolina (Clotfelter et al., 2008; Hoxby, 2006).

²¹The LUSD generated separate waitlists for students with and without siblings at the school. After the initial lottery offers and responses were processed, any seats not accepted were offered to students on these waitlists in an alternating fashion. Specifically, the first seat was offered to a student on the sibling waitlist, then the next was offered from the non-sibling waitlist, the third was from the sibling waitlist, etc.

²²Recall that magnet schools are purposefully built in particularly high-minority-share neighborhoods.

(Billings et al., 2014) or half the size of court-ordered desegregation in the '60s and '70s (Rossell and Armor, 1996; Guryan, 2004).

1.3.2 No Child Left Behind

In 2002, the No Child Left Behind (NCLB) Act was signed into law as an update to the Elementary and Secondary Education Act of 1965. Because NCLB and race-blind lotteries were contemporaneously implemented, NCLB accountability measures present potential concerns for the validity of my identification strategy. In this section, I provide details about how this district implemented NCLB. I reserve discussing how NCLB may threaten the validity of my estimation strategy for section 1.6.3 to allow the discussion to occur in the context of my empirical method.

One of the earliest consequences for a school that fails to meet NCLB-determined academic requirements is to be subjected to increased competitive pressures through school choice. Starting in the 2003-04 school year, the LUSD required every school in the district (including magnet schools) to set aside a portion of their seats for the NCLB placement mechanism.²³

Students across the district were ranked using two inputs: the student's baseline testing and family income, where a low ordinal ranking signified the lowest achieving, poorest students in the district. Students currently assigned to a traditional school in the district that failed to meet NCLB-determined academic measures were eligible to participate in NCLB school placement. Prior to the magnet school admissions lotteries, students from these failing schools would rank order up to three schools in the district into which they wanted to transfer. The student with the lowest rank (most disadvantaged) was placed first, followed by the next lowest ranked student, and so on. If the student's first-choice school had no more NCLB seats, then the student would be placed in their second-, or third-choice school. If all three choices were full, the student would not receive a NCLB-seat and would have to apply to the magnet school lotteries as before. After NCLB seats were determined, the (now race-blind) magnet school lotteries were carried out normally as explained in Section 1.3.1.

²³LUSD magnet middle schools set aside roughly 20% of their 6th grade seats for NCLB placements.

1.4 Data

I use student-level administrative data from a large urban school district (LUSD) from 1998 through 2007. As a condition to access their data the district requested complete anonymity. This district enrolls roughly 40 to 60 thousand students in traditional schools and 10 to 15 thousand students in magnet schools in any given school year.

In addition to statewide standardized achievement measures and student demographic information, the district also merged student information to several medium- to long-run student outcomes.²⁴ The district matched student records with ACT/SAT achievement from 2004 through 2011 and also merged student records for each graduating class with college information collected by the National Student Clearinghouse (NSC).²⁵ NSC data include the name of each college attended and the student’s major as well as whether and when they graduated from college. The NSC covers all public and private, two- and four-year postsecondary institutions in the United States allowing me to observe students attending out-of-state schools.²⁶ The LUSD combined these student-level data with admissions lottery records over the same time horizon which includes information on which schools each student applied in a given year and any seat offers. From waitlist information, I can deduce which students were offered seats during the initial wave, hereafter denoted “initial offers.” I am also able to observe basic demographic information for all teachers in the district and, for 2000 and later, I can link students to their teachers and classmates.

Prior to any sample restrictions, I observe roughly 50,000 6th grade students from 1998 to 2007.²⁷ Data contain students attending any of the traditional or magnet public schools within the LUSD, thus I cannot observe any students who transfer to a charter or private school or who move out of

²⁴Student demographic information is only available from 2000 onward, but I infer student race for earlier years based on which race-specific lottery they utilize.

²⁵I am in the process of matching NSC records to all lottery applicants to avoid any differential attrition concerns.

²⁶See [Dynarski et al. \(2013\)](#) for further details on NSC coverage rates across institution types.

²⁷Starting in 2008-09, the LUSD set aside a third of the seats within several of the most popular magnet schools for a separate selective-admissions lottery. Students who were categorized as “Gifted” or who tested in the top 5 percent of the district on a standardized test in 6th grade were eligible to apply using this smaller lottery. Because eligible students were disproportionately white, the share of white students in magnet schools rapidly increased starting in the 2008-09 school year. As a result, for this paper, I restrict attention to lotteries occurring prior to 2008-09.

the area entirely. Table 2.1 presents descriptive information about the student composition of this LUSD. Column 1 shows that for the full sample, the district is composed almost entirely of black and white students (cumulatively 92%) with a majority of the district being comprised of black students. Because race-specific admissions lotteries were conducted separately for black and non-black students, I similarly consider students of other races and ethnicities as non-black throughout the paper.

On average, black students in this district test below non-black students. Figure 1.2 displays the distribution of scores among black and non-black students in the district broken out by subject. The distribution of black test scores lies to the left of the non-black distribution for all subjects.

I restrict the sample to students who have applied to at least one magnet school in 6th grade and do not come from a sending school with automatic placement in a magnet middle school. The sample is further restricted to students without sibling priority in any magnet lottery. I also exclude lotteries from the 2001-02 school year because observable student characteristics fail to balance across lottery winners and losers for this year. Finally, given these restrictions, I drop any students who are the only ones in the district applying to the given magnet lottery after other sample restrictions are applied.

Column 3 of Table 2.1, shows descriptive information for all students in this baseline estimation sample. The sample for this table further requires that students have valid reading achievement outcome information. These conditions limit the sample to roughly 6,000 student-year observations for 6th grade applicants. Students in this lottery sample are more likely to be female and black. Additionally, students in the magnet school regression sample have higher baseline achievement performance across all four subjects.

1.5 Estimating the Baseline Returns to Magnet School Attendance

1.5.1 Magnet School Lottery

As a baseline, I establish how magnet schools compare to traditional schools by estimating differences in school inputs, peer composition, student achievement, and long-run outcomes for students who win a magnet school lottery seat relative to those who do not. Specifically, I estimate

$$y_{il} = \rho M_{il} + \Gamma_{2l} + \gamma' X_i + \epsilon_{il} \quad (1.1)$$

where y_{il} is an outcome for a student i who applies to the 6th grade magnet school lottery l .²⁸ X_i is a vector of pre-lottery demographics that includes indicator variables for student race (black or non-black) and gender. Similar to Billings et al. (2014), X_i also includes quadratics in pre-lottery baseline reading, math, science, and writing achievement as well as missing achievement indicators for each subject. M_{il} is an indicator equal to one if the student enrolled in a magnet school during the year following the lottery.²⁹ Γ_{2l} are lottery indicators, i.e., a unique application-school-by-lottery-type-by-year combination.³⁰ Because the unit of observation is a student-application, standard errors are two-way clustered by student and the enrolled school after the lottery in 6th grade. Further, regressions are weighted by the inverse of the number of applications submitted by the given student so that each student contributes equally to the regression.

If magnet enrollment were randomly assigned, then ρ would give the causal effect of attending a magnet school in sixth grade on the given outcome. However, any unobserved determinants of student outcomes that also correlate with the decision to enroll in a magnet school would bias my estimate of ρ . The existence of such unobservable correlates seems likely given that magnet school applicants have higher baseline standardized test scores and are more likely to be black than other traditional public school students as shown in Table 2.1. As a result, I instrument for magnet enrollment using exogenous lottery offers through the following first-stage:

$$M_{il} = \Gamma_{1l} + \beta' X_i + \pi Z_{il} + \eta_{il} , \quad (1.2)$$

²⁸Note that if a student applies to multiple 6th grade lotteries the outcome is common across all lotteries.

²⁹Students are counted as being enrolled in a magnet school if they are enrolled for one or more days.

³⁰Lottery type refers to black, non-black, or race-blind lotteries.

where Z_{il} is an indicator variable equal to one if student i receives an initial magnet offer in lottery l . In a comparable estimation framework, Angrist et al. (2016) use both initial lottery offers as well as whether the student ever receives an offer as instruments to assess the returns to charter school enrollment. However, in my setting, because students do not rank their school preferences and once a student accepts a lottery offer they are automatically removed from all other waitlists, subsequent lottery offers from randomized waitlists are endogenous. To see this, suppose that wealthier families are more willing to wait for a magnet seat in their preferred school and that low-income families are more likely to accept the first school offer they receive. If this is the case, then while the set of initial offers should have an equal share of high- and low-income students offered a seat, there would be a disproportionately larger share of lottery offers that are *ever* extended to high-income families from the waitlist because they are more likely to have waited.

To ensure that lotteries only compare students with the same probability of receiving a magnet offer, all regressions condition on a full set of lottery effects Γ_l . Students share a lottery if during the same year they apply to enter the same magnet school in 6th grade through the same type of lottery (i.e., black, non-black, or consolidated race-blind lottery).

If offers are truly random, then predetermined student characteristics should be equally represented or “balanced” across winners and losers within lotteries. I test for lottery balance by regressing student observables on an indicator for whether the student receives a magnet offer to the given lottery’s reference school and a full set of lottery fixed effects. Column 5 of Table 2.1 presents these tests. Overall, lottery winners are comparable, on average, to losers across these observable dimensions. While not statistically different than zero, students with higher baseline reading test scores appear marginally less likely to win a seat. The combined p -value in the table is for a test of joint significance of the difference between lottery winners and losers across all outcomes. While this difference is statistically different at the 10 percent level, these individual offer differentials are comparable to other lottery studies in the literature (e.g., Abdulkadiroglu et al., 2011; Angrist et al., 2016). As a precaution, I include race and gender as well as baseline subject-specific achievement as controls throughout the paper. Overall, these regressions provide evidence that initial lottery offers are indeed random.

My empirical strategy is similar to Cullen et al. (2006), who estimate the reduced-form effect of

receiving a school lottery offer on achievement using application-level data. Cullen et al. (2006) are inherently interested in the effect of additional schooling options and so they focus on the direct effect of receiving a lottery offer on student outcomes. Because I am specifically interested in estimating how magnet enrollment impacts student outcomes, I instead pursue a two-stage least squares approach (2SLS) that provides the causal effect of *enrolling* in a magnet school among the set of students that are induced to enroll by the randomized lottery offers (Imbens and Angrist, 1994).

Further, it is important to emphasize that just as in Cullen et al. (2006) and Cullen and Jacob (2007) the unit of observation in my setting is a student-application. Thus, students who apply to multiple magnet schools will appear in the data multiple times.³¹ As a result, a student who wins one lottery and loses another will contribute to the treatment and control groups of the respective lotteries. Cullen and Jacob (2007) explain that this setup still produces consistent parameter estimates because randomization ensures that while some proportion of lottery winners also won seats in other lotteries, this is also the case among lottery losers. However, Cullen and Jacob (2007) go on to highlight that multiple applications do influence the magnitude of the treatment effect, because differences in outcomes between average lottery winners and losers will be more similar. An alternate strategy would be to employ a nested model that incorporates multiple magnet choices and student-year-level data similar to Angrist et al. (2016). However, subsequent segregation estimates (see Section 1.6) require the use of application-level data. Thus, to make baseline estimates more comparable with subsequent segregation estimates, I utilize the same application-level data in both settings.³²

1.6 Estimating Peer Racial Composition Effects

To ensure a pre-determined level of racial diversity in its magnet schools, this LUSD held separate lotteries for black and non-black students to fill seats in oversubscribed schools through the 2002-

³¹Students can apply to up to 3 magnet schools and, on average, students end up applying to 2 schools.

³²I also estimate baseline regressions using a framework comparable to Angrist et al. (2016) and find qualitatively similar results. These estimates are available upon request.

2003 school year. In subsequent years, this district instead used race-blind lotteries where the probability of winning the lottery was the same regardless of race or ethnicity. Figure 1.3 depicts how the introduction of race-blind lotteries impacted the probability of winning a magnet lottery each year by student race. Prior to 2003, because black families disproportionately applied to magnet schools, non-black students were 15 to 20 percentage points more likely to win an initial lottery seat than black students. The introduction of race-blind lotteries, denoted by the reference line in 2003, caused both black and non-black students to have nearly identical, albeit much lower probabilities of winning.

The large drop in win probability is due to the introduction of NCLB. In addition to consolidating the lotteries, in 2003-04, the LUSD implemented NCLB school choice requirements by setting aside seats in schools across the district for the least-proficient, lowest-income students from failing schools (see Section 1.3.2). Because these students are disproportionately black, the NCLB placement mechanism potentially further exacerbated racial imbalance within magnets. The drop in the probability of acceptance reflects the decrease in the number of seats available to be filled via lottery.

While the lottery regime change impacted the composition of the 2003-04 entering class, the adjustment did not directly affect magnet school curricula or teaching staff.³³ However, the concurrent passage of NCLB presents a possible confounder. Thus, simply comparing estimates of the effect of attending a magnet school before and after the lottery consolidation in 2003 would conflate any NCLB-driven impacts.

Prior to 2003-04, the racial composition of students receiving initial seat offers did not necessarily reflect the composition of the full applicant pool, but did so thereafter. To account for potential structural changes outside of the termination of race-based lotteries, I leverage the fact that the size of the shift in racial composition due to the lottery change varied by how “disproportionately non-black” that lottery offers were for each school. I measure how “disproportionately non-black” that offers were for a given school by calculating the difference between the percentage of black

³³Staff reshuffling could result from changes in student demographics (Jackson, 2009), however, none of these changes were structurally a part of the lottery consolidation and can be considered as part of the re-segregation treatment.

students in the lottery pool for the school to the percentage of black students receiving an initial magnet school offer during the 2002-03 school year (denoted DPB'^{02}).

This DPB'^{02} measure is useful because the larger the difference the larger the potential shift in school racial composition upon the lottery consolidation. To see this, consider a school (call it school A) where 80% of all 2002 lottery applicants were black, but due to the dual lottery system, the school offered only 50% of the seats to black students. Conversely, consider school B, where 50% of the students in the applicant pool were black and also that 50% of the students who received an initial offer were black. Supposing that the composition of the student applicant pool remains roughly the same from 2002 to 2003, after consolidation, the composition of black students offered a seat to school A would rise to 80% to mirror the applicant pool, while the racial composition of lottery offers to school B would remain unchanged.

I isolate the exogenous shift in school segregation due to the establishment of race-blind lotteries in an instrumental variables framework by estimating the following first-stage:

$$\begin{aligned} \%Black_{il} = & \rho DPB'_l{}^{02} * \mathbb{1}(\text{Post } '02)_t * \mathbb{1}(\text{Offer})_{il} \\ & + \kappa_1 \mathbb{1}(\text{Post } '02)_t * \mathbb{1}(\text{Offer})_{il} + \delta_1 DPB'_l{}^{02} * \mathbb{1}(\text{Offer})_{il} \\ & + \theta_1 \mathbb{1}(\text{Offer})_{il} + \gamma'_1 X_i + \Gamma_{1l} + \nu_{il} \end{aligned} \quad (1.3)$$

with the accompanying second-stage

$$\begin{aligned} y_{il} = & \beta \widehat{\%Black}_{il} \\ & + \kappa_2 \mathbb{1}(\text{Post } '02)_t * \mathbb{1}(\text{Offer})_{il} + \delta_2 DPB'_l{}^{02} * \mathbb{1}(\text{Offer})_{il} \\ & + \theta_2 \mathbb{1}(\text{Offer})_{il} + \gamma'_2 X_i + \Gamma_{2l} + \epsilon_{il} , \end{aligned} \quad (1.4)$$

where $\%Black_{il}$ is the leave-one-out percentage of black students enrolled in the school that student i attends during the year following the given lottery l .³⁴ DPB'^{02} is the 2002-03 application-school-specific difference in the percentage of black students in the lottery applicant pool relative to the percentage receiving an initial offer for the application school in lottery l . Specifically, $DPB'^{02} = 100 * \left(\frac{\sum_{i \in j} \mathbb{1}(\text{Black})_i}{N_j} - \frac{\sum_{i \in j} \mathbb{1}(\text{Black})_i \cdot \mathbb{1}(\text{Offer})_i}{\sum_{i \in j} \mathbb{1}(\text{Offer})_i} \right)$, where N_j is the total number of applicants

³⁴The leave-one-out percentage is calculated by ignoring the reference student and calculating the given statistic for the remaining 6th grade students in the school.

to school j .³⁵ $\mathbb{1}(\text{Post '02})$ and $\mathbb{1}(\text{Offer})$ are indicator variables respectively equal to one if the current lottery occurs strictly after the 2002-03 school year or if the student receives an initial seat offer in lottery l .³⁶ X_i and Γ_l are respectively the same set of pre-lottery characteristics and lottery-specific fixed effects from equation (1.1). Similarly, standard errors are again two-way clustered by student and school-after-lottery and regressions are weighted by one over the number of applications submitted by the given student in the given year so that each student equally contributes to the estimation.

I instrument for the percentage of black students using the triple interaction between my measure of lottery racial imbalance (DPB'^{02}), an indicator for whether the lottery occurred after 2002, and an indicator for whether the student received an initial magnet offer. To understand the interpretation of the coefficient ρ , consider two schools A and B where the 2002 lottery racial imbalance in A is one percentage point larger than B. ρ provides the differential effect of winning an initial seat after 2002 in the lottery for school A relative to B on the percentage of black students that will enroll in the student's 6th grade school. In other words, this instrument isolates the variation in racial composition that is induced by the change in the lottery regime across schools with differing levels of underlying lottery racial disparity. The identification assumption is that unobserved determinants of magnet school effects are not trending differentially by DPB'^{02} . For example, suppose that the schools in the neighborhoods experiencing "white flight" are also steadily declining in their effectiveness. If "white flight" is trending upwards in neighborhoods where high- DPB'^{02} schools are located, then trends in "white flight" and school effectiveness would bias my estimates.³⁷ However, in Section 1.6.1, I provide event studies that show little evidence for differential trends in school composition and productivity across school DPB'^{02} values.

This empirical strategy is able to account for a variety of potential confounders. First, I can handle changes to policies that are contemporaneous with the lottery consolidation. As long as other policy changes do not differentially affect schools by DPB'^{02} then these potential confounders

³⁵Several magnet schools have DPB'^{02} values near 0, while others have values ranging up to a 10 percentage point difference.

³⁶ $\mathbb{1}(\text{Post '02})$ and DPB'^{02} main effects are absorbed by lottery effects.

³⁷Gamoran and An (2016) find that upon the termination of court-ordered desegregation in Nashville, academically selective magnet schools became more white and non-selective magnets more black.

will be controlled for directly by the interaction between the initial offer and post-2002 binaries. Further, suppose that magnet schools that have higher lottery racial disparity are generally better schools. Then, the interaction between DPB'^{02} and the initial offer binary controls for this directly as long as the effectiveness of these schools do not change after 2002.

It is important to note that while I am estimating the effect of a plausibly exogenous racial composition shock, I will be unable to disentangle any other composition changes happening simultaneously. For example, because black students in this district test lower on average than non-black students, an exogenous increase in the percentage of black students at the school will likely decrease the average baseline standardized achievement as well. Thus, if peer achievement is the actual mechanism that affects own achievement, an ability peer effect would appear like a race peer effect. However, this is still an interesting parameter to estimate. Policy-makers aiming to increase racial diversity in schools are simultaneously changing not only racial make-up, but also socioeconomic status, aptitude, and an array of other student and teacher demographics (Jackson, 2009). As a result, while I am unable to isolate the effect of racial diversity on student outcomes, per se, I can estimate parameters relevant to real-world desegregation policies.

1.6.1 Validating the Instrumental Variables Strategy

In this section, I test whether the causal effect of winning a magnet seat trends differentially by my measure of pre-consolidation lottery racial disparity. I assess the power of my first stage by regressing the reduced form analog of equation (1.3) where I interact initial offers and lottery racial disparity with year indicators instead of a post-2002 binary. Specifically, I estimate

$$\begin{aligned} \%Black_{ijl} = & \sum_{\substack{t = 1998; \\ t \neq 2002}}^{2007} \left\{ \rho_t DPB'_l{}^{02} \cdot \mathbb{1}(\text{Year} = t)_t \cdot \mathbb{1}(\text{Offer})_{il} + \kappa_t \mathbb{1}(\text{Year} = t)_t \cdot \mathbb{1}(\text{Offer})_{il} \right\} \\ & + \delta_1 DPB'_l{}^{02} \cdot \mathbb{1}(\text{Offer})_{ijl} + \theta_1 \mathbb{1}(\text{Offer})_{il} + \gamma'_1 X_{il} + \Gamma_l + \nu_{ijl} \end{aligned} \quad (1.5)$$

where variable definitions are analogous to equation (1.3). Estimates are relative to 2002, the year prior to the lottery consolidation. If my empirical strategy successfully isolates the variation in racial composition driven by the lottery consolidation, then, relative to 2002, the effect of winning a seat in a more racially disparate 2002 lottery pool on the racial composition within the student's

enrolled school should be zero for 2001 and earlier and positive thereafter. Figure 1.4 displays estimates of ρ_t for this regression.³⁸ Indeed, prior to the consolidation, aside from 1999, winning a seat to a magnet school with a larger DPB'^{02} value has no statistically distinguishable effect on the percentage of black peers that eventually attend the lottery winner’s school of enrollment. However, upon the termination of race-conscious lotteries in 2003, I estimate that winning a seat to a school with a one percentage point larger DPB'^{02} value increases the proportion of black peers attending the school where the lottery winner enrolls by roughly 1.8 percentage points. Because the effect of magnet seat offers on the enrolled school racial composition does not systemically differ by DPB'^{02} prior to the consolidation, this is evidence for the absence of trends in unobservables that correlate with DPB'^{02} and also drive school racial composition.

Figure 1.5, presents estimates of the same regression, but for several important dimensions of school composition and student achievement. Panels 1.5a, 1.5b, and 1.5c respectively present estimates for the leave-one-out averages of baseline math and reading achievement and free/reduced lunch eligibility among peer 6th graders within the student’s enrolled school.³⁹ As expected, peer baseline math and reading scores drop after the lottery consolidation, while free-lunch eligibility increases. Because black students in this district have lower baseline achievement on average and a higher proportion are free-lunch eligible, it is not surprising that an exogenous increase in the proportion of black students within a grade affects student composition along these dimensions.

Panels 1.5d through 1.5f present event studies for several student outcomes. Figures for the remaining outcomes explored in this paper can be found in Appendix A. To concisely summarize outcomes, I create indices by respectively taking averages over standardized versions of the achievement outcomes (i.e., math, reading, science, and writing achievement) and the postsecondary attainment outcomes (i.e., college enrollment, 2-year, 4-year, and “Top 50” rank enrollment). Foreshadowing future results, student achievement and postsecondary attainment is negatively affected by winning a seat to a school with a higher DPB'^{02} value after 2002, while ACT test taking is unaffected. In general, the absence of pre-trends across these regressions supports the identifying assumptions underlying my estimation strategy.

³⁸Recall from section 3.4 that I exclude lotteries from 2001 because they fail to balance on observable student characteristics.

³⁹Free/Reduced price lunch eligibility comes from school-level averages from the CCD.

1.6.2 Differential Attrition

After testing whether the lotteries balance across observable student demographics, the other primary concern that potentially invalidates the lottery empirical strategy is differential attrition from the analysis sample between winners and losers. Suppose that wealthier families who lose a magnet lottery are more likely to send their child to a private school. Because I cannot observe students in private schools, this attrition would cause the lottery losers with valid outcome data to be disproportionately lower income, invalidating the empirical design.

In Table 1.2, I test whether lottery winners are less likely to be: enrolled within the district during the year following the lottery, missing math achievement outcome information, and missing NSC data.⁴⁰ In Panel A, I test differential attrition for my baseline estimates by regressing each attrition outcome on an indicator equal to one if the student was awarded an initial offer to the lottery's magnet school as well as a full set of lottery fixed effects. Lottery winners are indeed less likely to be missing from my analysis sample across all outcomes though the difference is only statistically significant for missing NSC outcome information. Magnet winners are about 2 percentage points less likely to be missing NSC outcome data.⁴¹ In Panel B, I test whether differential attrition presents a threat to my segregation estimates from equation (1.3) by further interacting the initial offer variable with a post-2002 indicator and with DPB^{02} . In order to be an issue, rates of differential attrition must vary by DPB^{02} levels and must shift after 2002. I find no evidence that the rates of differential attrition systematically change after 2002 across lotteries with varying DPB^{02} levels. Thus, differential attrition does not present a concern for my estimates of the effect of racial composition on student outcomes.

⁴⁰The virtue of NSC data is that students can be matched even if they leave the sample. However, because the LUSD only matched NSC data for graduating cohorts, differential attrition is still a concern in my setting. I am in the process of matching NSC data to all students in the lottery sample.

⁴¹The magnitude of differential attrition is comparable to the differential attrition estimated by Cullen et al. (2006) in the open enrollment system in Chicago.

1.6.3 Accounting for No Child Left Behind and Changes in the Composition of Applicants

Both the introduction of NCLB and the lottery consolidation in 2003-04 could plausibly alter the composition of the pool of magnet lottery applicants. NCLB requires all schools in the district to reserve seats for the poorest and least proficient students enrolled in failing schools. Further, students who utilize the NCLB placement mechanism do not apply to magnet lotteries. As a result, one might expect the lottery pool to include students with higher baseline achievement and family resources than before NCLB. The lottery consolidation itself may also influence the composition of the lottery pool. Recall that the lottery consolidation caused the probability of winning a seat for non-black relative to black students to fall. Thus, one might expect to see a higher share of black students in the lottery pool after the consolidation.

While randomization ensures that average lottery winners and losers are comparable along observable and unobservable dimensions, changes in the composition of the lottery pool will impact how to interpret the treatment effect. To see this, suppose that NCLB increases the baseline achievement within the lottery pool as explained above. Even though the composition of the lottery has changed, lottery offers are still randomized among the new pool. Thus, the average baseline characteristics of winners and losers will be indistinguishable, but now both average lottery winners and losers have a higher baseline aptitude. This is only a problem if the effects of magnet enrollment are heterogeneous along the same dimension.

Suppose that magnet schools relative to traditional schools are better equipped to teach high-aptitude than low-aptitude students and that the effect of attending a magnet school is constant over time. Under these assumptions, because the lottery pool is filled with higher achieving students after 2002 and magnets are better at instructing these students, then the average returns to winning a magnet lottery would be larger after 2003 than before. However, this change simply reflects the shift in the composition of the lottery pool and not any change in the underlying education production of magnet schools over time.

This consideration does not pose a problem for my baseline estimates in Section 1.7, but potentially threatens how I identify segregation effects in Section 1.8. Specifically, if treatment effects are het-

erogeneous and if the composition of lottery applicants changes before and after 2002 differentially based on how disproportionately non-black the school’s lottery offers were in 2002 (DPB'^{02}), this would bias my estimated impact of segregation.

I test for compositional changes in lottery applicant pools across DPB'^{02} by regressing

$$y_{jt} = \beta DPB'_j{}^{02} * \mathbb{1}(\text{Post } 2002)_t + \gamma_t + \theta_j + \epsilon_{jt}, \quad (1.6)$$

where y_{jt} is the average of a pre-lottery characteristic across students applying to enter magnet school j in year t .⁴² DPB'^{02} is defined as in (1.3), $\mathbb{1}(\text{Post } 2002)$ is an indicator variable equal to one if the lottery occurs after the lottery consolidation in 2002, while γ_t and θ_j are respectively year and school effects.⁴³ Consider two schools A and B, where school A has a one percentage point larger DPB'^{02} value than school B. β provides the average difference in how the lottery pool for school A changes after 2002 relative to how school B changes for the given student characteristic.

Table 1.3 depicts estimates of equation (1.6) for student race, gender, and baseline achievement. I find that the proportion of black students applying to schools with a one percentage point higher DPB'^{02} value increases by 0.8 percentage points. I also find that the proportion of female students decreases by 0.3 percentage points. The largest DPB'^{02} value in the district is roughly 10, meaning that terminating race-based lotteries shifts the share of black and female students in the lottery pool for the school with the largest 2002 lottery racial disparity upwards by 8 percentage points and downwards by 3 percentage points, respectively. Considering that on average, the proportion of black students applying to a magnet school is roughly 0.80 and the proportion of female applying is 0.55, I consider these compositional changes as second-order concerns.

Additionally, I find no statistically significant shifts in baseline achievement. Because the lottery composition is only changing with respect to race and gender, estimating the effect of segregation separately by these two groups will eliminate these compositional effects. In order for this to be successful, it needs to be the case that conditional on race/gender the composition of students is fixed. In Panel B, I test for changes in achievement within race and gender categories by restricting the sample appropriately. Indeed, I am unable to detect significant changes to the baseline achievement of the lottery applicants within these groups. Thus, I can abstract from

⁴²The regression is weighted by the number of students applying to the given school in the given year.

⁴³Main effects for $\mathbb{1}(\text{Post } 2002)$ and DPB'^{02} are absorbed by these indicator variables.

compositional changes to the lottery pool by simply exploring sub-group analyses of segregation effects.

Aside from these composition issues, NCLB also potentially changes how to interpret the treatment. Because black students have lower average achievement than non-black students in the district (see Figure 1.2), NCLB seats may be disproportionately awarded to black students. I consider DPB'^{02} to proxy for the underlying black student demand for the given magnet school. Thus, if a higher proportion of black students fill NCLB seats for schools with higher DPB'^{02} values, then the lottery consolidation will appear to induce additional racial imbalance into the school. While this poses no threat to internal validity, it impacts how to interpret the treatment. In Table 1.4, I show that the composition of students awarded NCLB seats to magnet schools does not statistically significantly differ by the school's value of DPB'^{02} .⁴⁴ As a result, I interpret the results from equation (1.3) as isolating the exogenous change in racial composition solely resulting from the termination of race-blind lotteries.

1.7 Magnet Enrollment Effects

1.7.1 Effects of Magnet Enrollment on Teacher and Peer Characteristics

Before estimating the impact of racial segregation on magnet school effects, I first benchmark the returns to magnet enrollment within the LUSD. I begin by analyzing how magnet school enrollment changes a student's exposure to different dimensions of teacher quality, school institutional details, and peer composition. Together these effects help characterize the magnet enrollment treatment, which will be useful in determining whether the racial composition effects I estimate in magnet schools may generalize to traditional schools as well. Table 1.5 presents the effect of magnet enrollment on a variety of teacher and peer characteristics from equation (1.1). Recall that the endogenous variable of interest is an indicator variable equal to one if the student enrolled in a magnet school during 6th grade. I instrument magnet enrollment with an indicator equal to one if

⁴⁴Specifically, I regress each outcome on DPB'^{02} values and year indicators among students who accept NCLB-provided seats to magnet schools in 6th grade.

the student won a seat in the given magnet school lottery during the first wave of offers.

Panel A displays two-stage least squares (2SLS) estimates along with the accompanying first stage estimates for the pooled sample. The first stage estimates characterize the take-up rate of the lottery offer treatment. There are many reasons why magnet school enrollment may not perfectly correlate with initial offers. First, anyone who receives an initial magnet school offer still has the prerogative to enroll elsewhere. Further, students who do not receive an initial lottery offer may eventually receive a seat after being waitlisted, letting them gain entry to the magnet despite losing the initial lottery. Depending on the estimate, students receiving an initial middle school magnet offer are anywhere from 15 to 19 percentage points more likely to be enrolled in a magnet school. These take-up rates are comparable to other lottery studies in the literature.⁴⁵

In columns 1 through 3, I assess the differences between magnet and traditional schools along several classroom measures. Specifically, among the set of classrooms a given student takes in 6th grade, I calculate the average teacher experience and class size as well as the proportion of teachers with a Masters degree that the student is exposed to throughout the year. Students that attend magnet schools are assigned to classrooms where the teachers are no more likely to have a Masters degree. However, students in magnet schools are taught by less-experienced teachers and attend classes that are on average about 4 students larger than in traditional public schools. Panel B displays 2SLS estimates separately for black and non-black student subgroups. Interestingly, non-black students who attend magnet schools are exposed to a higher fraction of teachers with Masters degrees, but who are less experienced on average.

Because classroom-specific information is only available from 2000 and later, in columns 4 through 7, I estimate the effect of magnet attendance on peer composition at the school level (as opposed to the classroom level) to exploit a larger sample more in line with subsequent analyses. These outcomes are school-year-grade-specific averages of peer compositions that omit the student's own characteristic. From the perspective of non-black students, enrollment in a magnet school increases the proportion of black students in the cohort. Conversely, for black students, magnet enrollment decreases the share of students in the cohort qualifying for free/reduced lunch (FRL) and increases

⁴⁵For example, Angrist et al. (2016) estimate that winning an initial lottery seat to attend a Boston charter school increases subsequent charter enrollment by 15 to 22 percentage points.

peer baseline academic achievement across both reading and math. In general, exposure to magnet schools in this LUSD affects teacher- and school-level educational inputs, but the imprecise estimates prevent me from ruling out relatively large differences in the composition of peers between magnet and traditional schools.

In summary, the effect of attending a magnet school is an amalgamation of school practices, teacher characteristics, and substantial changes to the composition of a student’s peer group. Each present possible mechanisms driving the effects estimated in the following sections. While I am unable to isolate the role of school and teacher inputs, in Section 1.8, I exploit a natural experiment that isolates the effect of changing the peer racial composition on magnet returns.

1.7.2 Effects of Magnet Enrollment on Student Outcomes

In this section, I test whether magnet school enrollment impacts student outcomes relative to traditional schools in this LUSD. The value of this analysis in the context of studying the effect of segregation on student outcomes is twofold. First, because I am exploring how magnet school productivity changes due to increased racial segregation, baseline estimates for magnet school productivity relative to traditional schools are useful to interpret subsequent segregation effects. Second, these estimates help inform whether segregation effects are externally valid. If the returns to attending a magnet school are similar to that of traditional schools in this district, then the segregation effects that I find in the magnet school setting may more plausibly generalize to traditional schools as well.

Table 1.6 presents the effect of magnet enrollment on several student outcome summary measures. Adapting the method used by Billings et al. (2014), I create indices that summarize student achievement in column 1, postsecondary attainment in column 2, and total student academic outcomes in column 3. The achievement index is the simple average across student middle school math, reading, science, and writing achievement.⁴⁶ The postsecondary index is a simple average over standardized versions of whether the student enrolled in any postsecondary institution as well as a 2-year, 4-year,

⁴⁶If a student is missing outcome information for a subject, then the average is taken over the remaining subjects only.

or “Top 50” ranked institution.⁴⁷ In addition to these outcomes, the total index also averages over standardized versions of high school graduation status, ACT test taking status, and ACT composite scores. Because the focus of this article is assessing how segregation influences education production, I relegate estimates of magnet effects on individual outcomes and the accompanying discussion to Appendix B.

Panel A provides estimates for the effect of magnet enrollment on student outcomes among the pooled regression sample. I am unable to detect statistically significant differences between magnet and traditional enrollees across all three indices though I am unable to rule out relatively large effects across each outcome. Panel B presents these estimates among sub-groups by student race, gender, and whether baseline math achievement is above or below the district’s median.

Magnet attendance improves achievement respectively by .10 and .15 standard deviations for black and non-black students as well as male and female students though the effect is not statistically significantly different than zero for most estimates. The achievement gains to magnet enrollment are primarily driven by improvement in science (see Appendix B). It is worth emphasizing that magnet school enrollment conditional on being offered a seat varies dramatically for black and non-black students. Table C.1 shows that winning the magnet lottery increases subsequent magnet enrollment by roughly 30 and 17 percentage points for non-black and black students, respectively. Anecdotally, magnet schools in this district were historically marketed to non-black students as a way to prevent “white flight” to the suburbs, which could help explain the difference in acceptance rates. As a result, the statistical power of the instrument also varies by student subgroup (e.g., Kleibergen-Papp F statistics for tests of weak instruments range from 3 to 12 depending on the specification – see Table C.3). Thus, caution should be given to interpreting under-powered subgroup estimates.

I estimate that enrolling in a magnet middle school tends to decrease (increase) postsecondary outcomes high-aptitude (low-aptitude) students. Again, these estimates are imprecisely estimated and should only be considered as suggestive. Column 3 shows that across the outcomes explored in this paper, the returns to magnet schools relative to traditional schools are negligible in the pooled sample, with imprecise heterogeneous returns to certain subgroups.

⁴⁷Postsecondary outcomes are based on enrollment decisions made within 18 months after high school graduation.

Together, these results build upon the findings of the previous section. In general, magnet schools do not generate educational benefits to students over other traditional public schools in the district. However, magnet schools generate both positive and negative heterogeneous effects across student subgroups. The localized and seemingly contradictory nature of these effects could be driven by differences in magnet-school-specific teaching strategies and specialties, where particularly (in)effective schools could be driving estimates for certain student subgroups. However, because the focus of this paper is estimating segregation effects, I leave the exploration of the heterogeneous impacts across magnet school types to future work. I read these results as suggestive evidence that magnet schools generate returns similar to the traditional schools in the district and I argue that studying the effect of racial segregation within these magnet schools may reasonably provide insight into how an increase in racial segregation would influence traditional public schools more generally. With these baseline estimates in hand, I now turn to the focus of this study by estimating the effect of a sharp increase in school segregation on student outcomes.

1.8 The Effect of Segregation on Student Outcomes

In this section, I explore one of the fundamental assumptions underlying the ruling of the landmark 1954 case *Brown v. Board of Education* in that school racial segregation negatively impacts minority student outcomes. I test this assumption by providing a causal estimate of how the peer racial composition parameter enters into the magnet school education production function.

Table 1.7 presents instrumental variables estimates in Panel A from equation (1.4) of the effect of a one percentage point increase in the share of black peers at a student’s school on the composition of other peer characteristics. Recall from Table 2.1 that roughly 80 percent of magnet school enrollment is comprised of black students. As a result, an increase in the share of black students attending magnet schools should be thought of as an increase in school segregation. Panel B provides first-stage estimates as well as F statistics for tests of weak instruments. The interpretation of the first stage estimate is that winning a seat after 2002 to a magnet school with a one percentage point larger disparity (DPB^{02}) increases the percentage of black peers in the student’s school by

roughly two percentage points.⁴⁸

Increasing the proportion of black students entering magnet schools also shifts the student composition along the dimensions of socioeconomic status and prior achievement. Segregation increases the share of free-lunch eligible students at a one-to-one ratio and decreases average peer baseline achievement in reading, math, and science. A ten percentage point increase in the share of black peers decreases average peer baseline achievement from 0.11 to 0.14 student-level standard deviations.

Middle school standardized testing provides an early measure to assess whether education production is sensitive to school racial composition. In Table 1.8, I estimate that a 10 percentage point increase in the share of black peers at a student’s school decreases achievement across math, science, and writing by 0.12 standard deviations. This is equivalent to the estimated achievement losses that would accompany permanently increasing class sizes by roughly 6 students (Angrist and Lavy, 1999).⁴⁹ The current literature estimates segregation-induced-losses to math achievement of 0.04 to 0.07 standard deviations, making my estimates somewhat larger (Hanushek et al., 2009; Billings et al., 2014). However, direct comparisons are obfuscated by the methodological differences and the unique educational setting in each study. My estimates are most closely aligned with Hoxby (2000) who estimates that a 10 percentage point increase in the share of black peers in a student’s class decreases math achievement by 0.19 points for black students.

Heterogeneous effects also confirm patterns in the segregation literature.⁵⁰ Math and science losses estimated among the sample of black students are even more pronounced, though statistically indistinguishable, from the pooled sample estimates. This suggests that achievement losses may be larger for black students than non-black students as in Hoxby (2000) and Hanushek et al. (2009). Decreases in math and science achievement are concentrated among black students with the highest baseline achievement as in Hanushek and Rivkin (2009), while segregation negatively impacts

⁴⁸The shift in the share of black peers is greater than unity because the lottery consolidation increased the proportion of black students applying to the magnet lotteries (see Table 1.3).

⁴⁹Angrist and Lavy (1999) estimate that a class reduction of 8 pupils increases reading achievement by about .18 student-level standard deviations.

⁵⁰I exclude non-black-student-specific results because I am severely under-powered due to the low counts of non-black students in the district.

writing achievement for low-achieving black students the most. Science and writing achievement losses are larger for black male students, while losses to math achievement for black female students are slightly more pronounced than for black male students.

Segregation has a clear negative impact on student achievement, but these short-term losses do not guarantee longer-term penalties to important education milestones such as high school graduation and postsecondary attainment. However, in Table 1.9, I show that the negative consequences of racial segregation are visible across several important medium-to-long-run educational outcomes. I estimate that a 10 percentage point increase in the share of a student's peers that are black erodes high school graduation rates by 4 percentage points. These losses are driven by female students whose high school graduation rates fall by twice the magnitude of the pooled sample.

School segregation also has deleterious impacts on postsecondary attainment. A 10 percentage point increase in school segregation decreases student enrollment in any postsecondary institution 18 months after high school graduation by 5 percentage points. The magnitude of this effect is comparable to exposing a student from 6th through 12th grade to teachers having nearly a standard deviation lower value-added (Chetty et al., 2014b).⁵¹ These losses are driven by low-aptitude black students as well as black female student subgroups and add to the myriad studies showing that the effects of educational interventions are often driven by particular subgroups, namely by minority and female students (Hastings et al., 2006; Deming et al., 2014; Anderson, 2008; Angrist et al., 2009; Angrist and Lavy, 2009; Jackson, 2013b; Andrews et al., 2016). Further, these heterogeneous effects are interesting because they remove any issues relating to the changing composition of magnet school applicants after 2002 (see Section 1.6.3).

Unlike in Charlotte-Mecklenburg where Billings et al. (2014) find that school segregation impacts four-year college attainment, I find that segregation instead discourages prospective college students from enrolling in 2-year institutions, particularly for black female students. This could be a direct result of the impact that segregation has on high school graduation rates. Students on the margin of graduating high school are more likely potential candidates to attend 2-year rather than 4-year

⁵¹Chetty et al. (2014b) estimate that one year of being exposed to a teacher with a one standard deviation lower value-added decreases college enrollment at age 20 by .82 percentage points. If you assume these effects accumulate linearly over time this yields a decrease of 5.74 percentage points for continual exposure from 6th through 12th grade.

postsecondary institutions. As a result, because segregation reduces high school graduation rates it is intuitive that I find that racial isolation subsequently diminishes postsecondary attainment through 2-year enrollment.

Finally, I assess whether college quality is affected by middle school segregation by estimating the impact on the probability of a student enrolling in a US News and World Report “Top 50” ranked university. While I estimate that segregation has no statistically significant impact on general college quality, I find that a ten percentage point increase in the share of black peers decreases student enrollment at a “Top 50” institution by 1 percentage point among black female students. Conversely, I estimate that black, male students are more likely to enroll in a “Top 50” institution by the same magnitude.

From these results, I conclude that racial balance is a meaningful input into the magnet school education production function both for immediate achievement as well as for long-run college outcomes. Magnet schools in this district generate similar returns to the traditional schools, and further, school-specific contributions to student achievement are intentionally removed by my estimation strategy. As a result, it is plausible that these results are not limited to the magnet school setting, rather, racial balance is likely an important input in education production functions more generally.

1.9 Conclusion

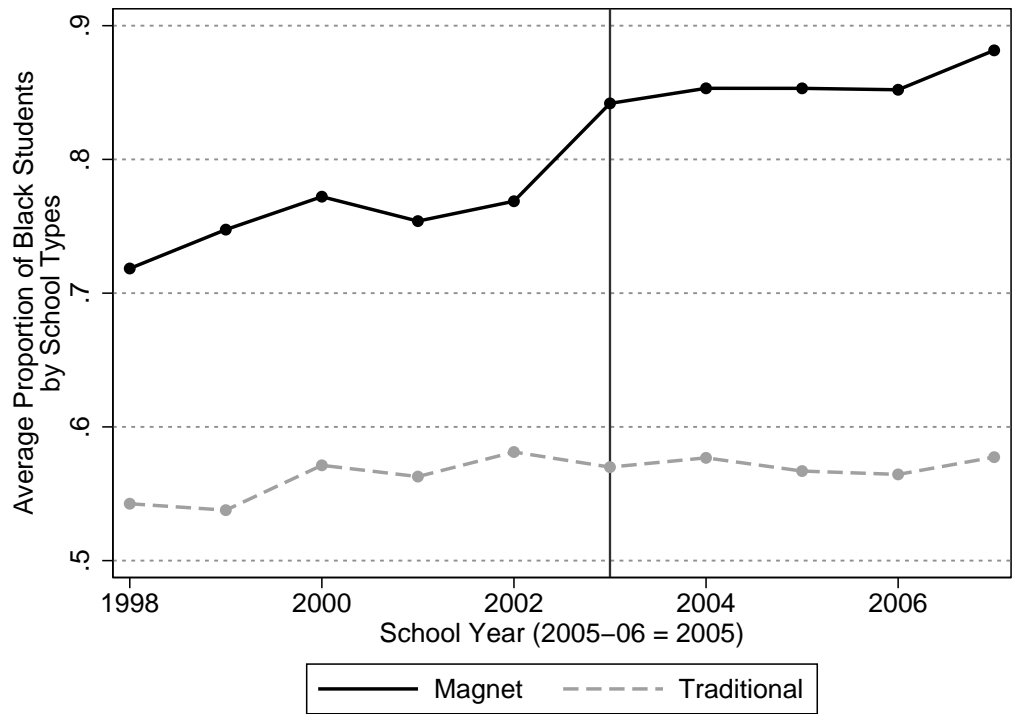
The United States education system has grown increasingly segregated since the end of court-ordered desegregation (Lutz, 2011; Reardon et al., 2012; GAO, 2016; Clotfelter et al., 2006, 2008), which makes assessing the causal link between *de facto* racial segregation and student achievement important to both researchers and policy-makers alike. I isolate how the effect of attending magnet schools in a large urban school district (LUSD) changes once schools are no longer allowed to artificially maintain the racial balance of incoming cohorts through race-based admissions lotteries. This setting provides an excellent natural experiment that reflects the growth in racial imbalance spreading across the nation’s school system.

I find that achievement gains to magnet enrollment are local to certain subjects and student subgroups. The localized nature of these returns leads me to conclude that, in general, the LUSD magnet schools have similar returns to other traditional public schools in this district. Further, my main estimates suggest that racial balance is an important input in the magnet school education production function. Exogenously increasing segregation decreases achievement among black high-achievers and negatively influences postsecondary outcomes among black females and black low-achievers.

While school assignment policies that explicitly use race in admissions decisions have been declared unconstitutional (*Parents Involved in Community Schools v. Seattle School District No. 1* – 2007), my results suggest that more creative policies aimed at improving racial balance in schools can generate large improvements in education production. For instance, many districts utilize information about residence instead of race to ensure their schools enroll a diverse student body from rich and poor neighborhoods. This work is particularly timely given President Obama’s current “Stronger Together” initiative that proposes to double the amount of federal funding up to \$120 million to support voluntary integration programs across the United States of which magnet schools are a part. Education interventions of any variety that incorporate achieving racial balance as either an explicit goal or simply a byproduct of the policy stand to improve student outcomes.

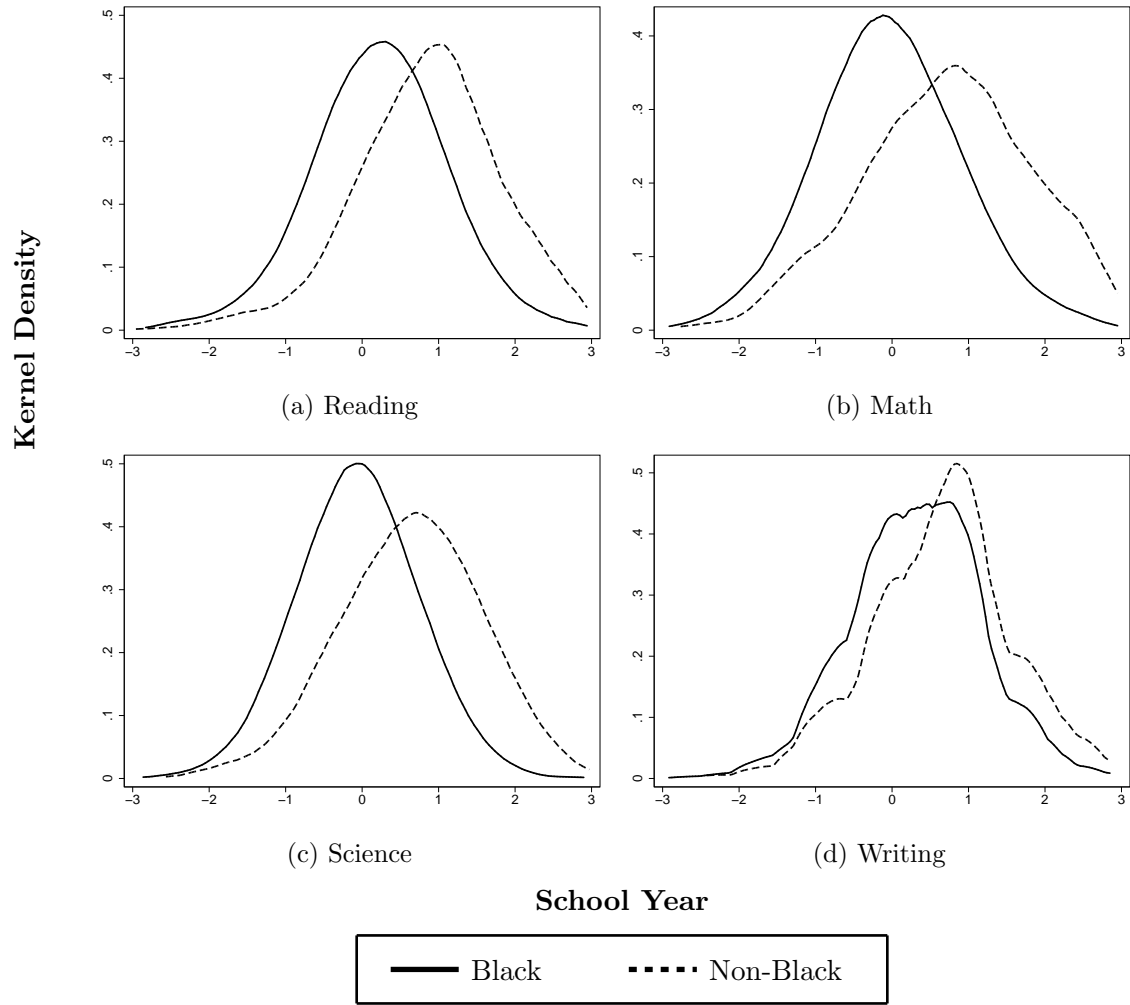
1.10 Figures and Tables

Figure 1.1: Racial Composition of Enrollment by School Type



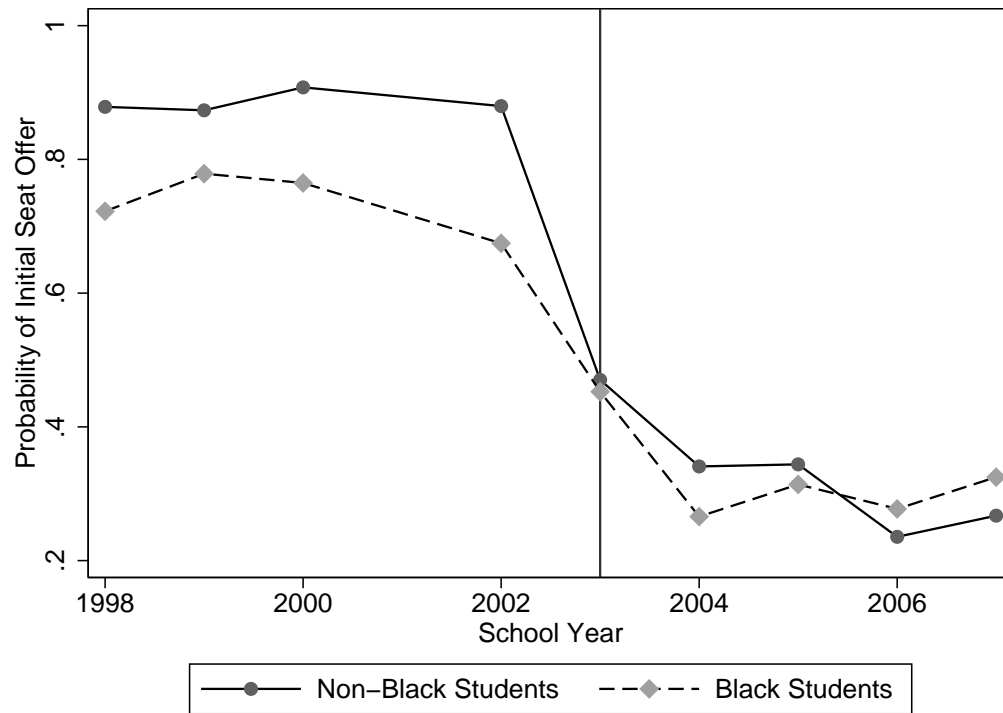
Notes: The figure plots the average black 6th grade enrollment shares across magnet and traditional schools in the LUSD. The vertical line represents the first year in which race-blind lotteries were used to determine enrollment in oversubscribed schools.

Figure 1.2: Standardized achievement Distribution by Subject and Race



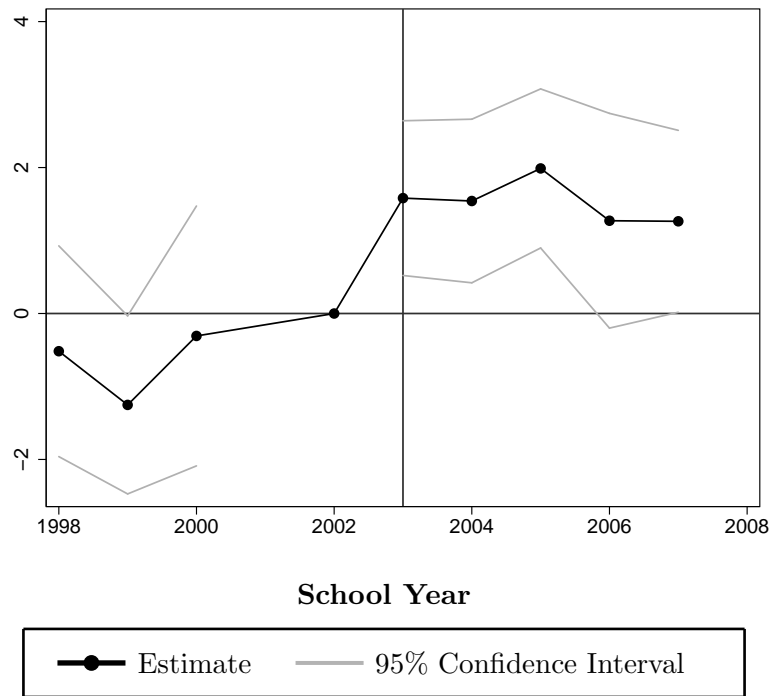
Notes: The figures present the distribution of subject-specific standardized achievement for 6th grade students who applied to magnet school lotteries prior to 2007.

Figure 1.3: Changes in Probability of Winning a Magnet Seat by Race



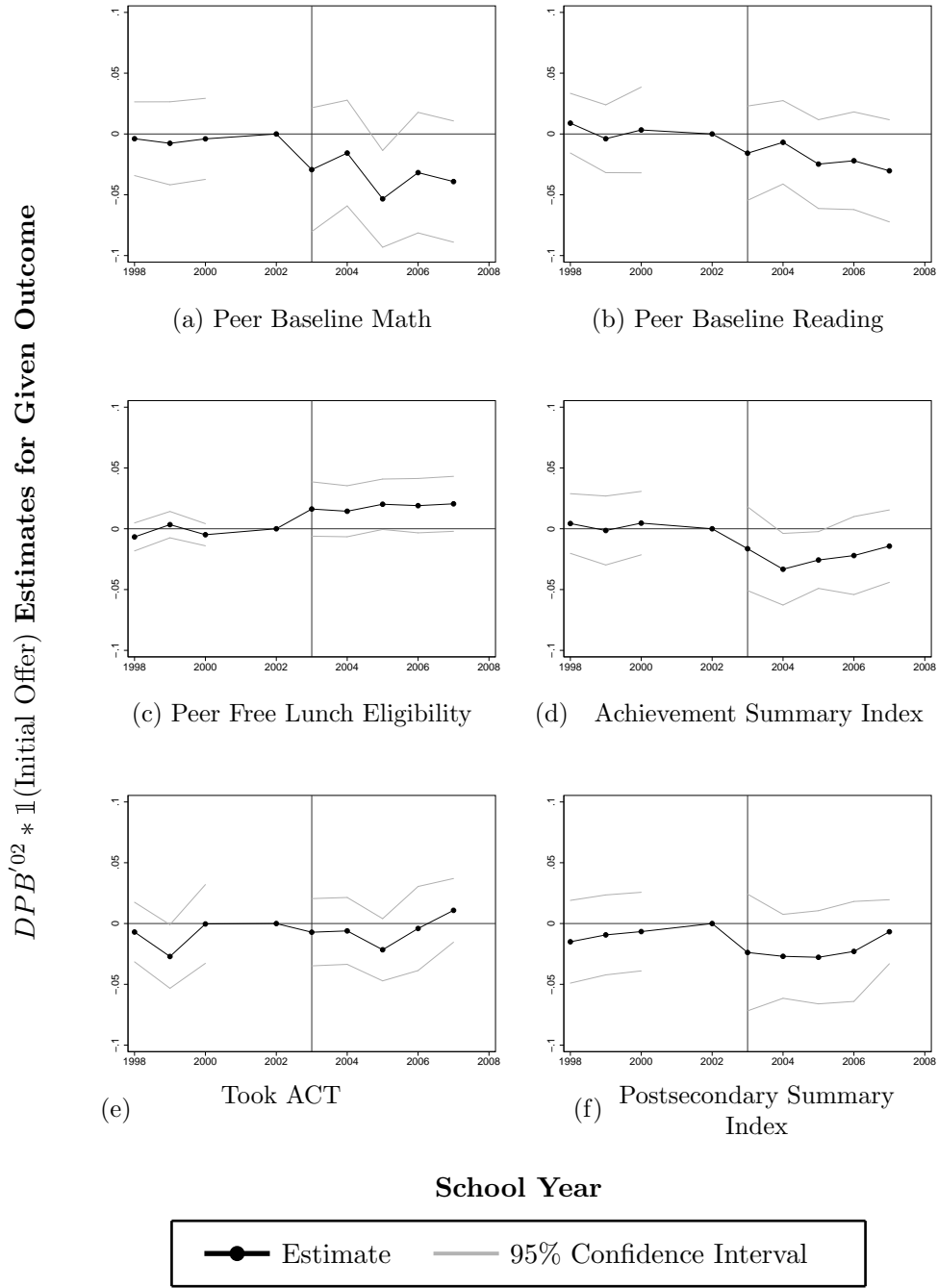
Notes: The figure plots the probability of winning an initial seat to attend a magnet school in 6th grade over time for by student race. The reference line in 2003 denotes the first year that the LUSD utilized the consolidated lottery to determine magnet seat offers. Prior to 2003, the LUSD ran separate lotteries for black students and non-black students (i.e. any student not considered a black student).

Figure 1.4: Trends in Black Student Composition by Lottery Racial Disparity (DPB'^{02})



Notes: The figure presents the effect of receiving an initial seat offer to a magnet school through a lottery with a 1 percentage point larger disparity between the percentage of black students in the lottery pool and the percentage of black student receiving offers in 2002 (DPB'^{02}) on the percentage of black students of the student's subsequent school of enrollment. Regressions are estimated using (1.5) as explained in Section 1.6.1. The regression is run using the sample restrictions in the footnote of Table 1.6. The reference line in 2003 denotes the first year the LUSD implemented the consolidated lottery system.

Figure 1.5: Trends in Various Outcomes by Lottery Racial Disparity (DPB^{02})



Notes: Each figure presents the effect of receiving an initial seat offer to a magnet school through a lottery with a 1 percentage point larger disparity between the percentage of black students in the lottery pool and the percentage of black student receiving offers in 2002 (DPB^{02}) on the given current outcome. Regressions are estimated using (1.5) as explained in Section 1.6.1. Each regression is respectively run using the sample restrictions for the given outcome in the footnotes of Tables 1.5 through 1.6. Similar Figures for the remaining outcomes explored in the paper are located in Appendix A. The reference line in 2003 denotes the first year the LUSD implemented the consolidated lottery system.

Table 1.1: Descriptive Statistics and Balance Test

	All LUSD Students		Magnet Lottery Sample		
	Mean (1)	N (2)	Mean (3)	N (4)	Initial Offer Gap (5)
Female	.482	53,833	.554	6,615	-0.005 (0.012)
Black	.620	52,915	.817	6,564	-0.009 (0.007)
White	.304	52,915	.155	6,564	0.011 (0.007)
Baseline Math	-.011	41,289	.282	5,127	0.001 (0.017)
Baseline Reading	.023	44,750	.231	5,205	-0.026 (0.019)
Baseline Science	-.000	36,753	.221	4,572	0.013 (0.017)
Baseline Writing	.018	31,353	.264	4,440	0.010 (0.016)
Combined p -value:			0.084		
Initial Offer Rate:			.514		
Magnet Attendance Rate:			.783		

Notes: The sample for columns 1 and 2 includes student-year observations for students attending a school 6th grade from 1998 to 2007. Columns 3 through 5 further restrict the sample to students who applied to a magnet school through the lottery and do not come from a sending school with automatic placement into a magnet school. The sample is further restricted to students without sibling priority in any lottery application and also to risk sets that have more than one student. Finally, the sample is further restricted to students with non-missing reading outcome test scores who are not receiving special education and are not repeating the given grade. Column 5 regresses each student demographic on an indicator equal to one if the student received an initial offer to a magnet school and a full set of lottery risk set fixed effects ($N = 13,315$). Standard errors (in parentheses) are clustered on students. p -values test the hypothesis that all coefficients on initial offer indicators are zero.

Table 1.2: Differential Attrition as a Function of DPB'^{02}

	Not Enrolled Following Lottery	Missing Math Outcomes	Missing NSC Outcomes
	(1)	(2)	(3)
Panel A: <i>Baseline Estimates</i>			
$\mathbb{1}(\text{Initial Offer})$	-0.003 (0.009)	-0.003 (0.007)	-0.020* (0.011)
Panel B: <i>Segregation Estimates</i>			
$DPB'^{02} * \mathbb{1}(\text{Post '02}) * \mathbb{1}(\text{Initial Offer})$	-0.007 (0.006)	-0.003 (0.003)	0.002 (0.006)

Notes: $N = 28,463$. The outcomes for each column are respectively a binary for whether the student appears in the LUSD enrollment records during the year following the 6th grade lottery, whether the student is missing math achievement information, and whether the student is missing in the National Student Clearinghouse data. Panel A presents the results from regressing the given outcome on a binary for whether the student was offered an initial magnet seat and a full set of lottery fixed effects. Panel B presents the results from a similar regression where instead of a binary for receiving an initial lottery offer, I include a triple interaction of the differential in 2002 percent black in the pool and winning a lottery (i.e., DPB), an indicator for if the lottery occurred after 2002, and the initial offer indicator, as well as the main effects. Standard errors for both regressions are clustered at the lottery level. Together these regressions show that there are no differential attrition problems in general (Panel A) and specifically for my racial segregation identification strategy (Panel B). Regressions are limited to observations from the baseline sample (see table note for Table 1.6).

Table 1.3: Differential Effect of Consolidation by DPB'02 on Middle School Applicant Composition

	Black (1)	Female (2)	Baseline Testing			
			Reading (3)	Math (4)	Science (5)	Writing (6)
<i>Panel A: Pooled Composition Changes</i>						
$DPB'_{02} * \mathbb{1}(\text{Post } 2002)$	0.008*** (0.002)	-0.003** (0.001)	0.004 (0.012)	-0.003 (0.012)	-0.004 (0.015)	-0.007 (0.008)
Observations	62	62	61	54	48	48
<i>Panel B: Baseline Achievement Composition by Race and Gender</i>						
$DPB'_{02} * \mathbb{1}(\text{Post } 2002)$	Black		0.008 (0.010)	0.003 (0.010)	0.003 (0.012)	-0.004 (0.007)
	Non-Black		0.009 (0.017)	0.004 (0.019)	-0.009 (0.022)	-0.007 (0.014)
	Female		0.006 (0.011)	0.002 (0.008)	-0.004 (0.013)	-0.004 (0.009)
	Male		0.003 (0.013)	-0.008 (0.016)	-0.005 (0.016)	-0.009 (0.008)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.6) where school-year averages of each outcome is regressed on an indicator equal to one if the lottery occurred after 2002 interacted with the school-specific DPB'_{02} value as well as indicators for the application school and lottery year. Regressions are weighted by the number of applicants in the given school-year pool. Standard errors (in parentheses) are clustered by application school. Each regression sample is limited to baseline sample restrictions specified in the notes for Table 1.6 who are also applying to enter a magnet school in the 6th grade.

Table 1.4: Composition of NCLB Magnet Seats as a Function of DPB'_{02}

	Black (1)	Female (2)	NCLB Rank (3)	Baseline Testing			
				Reading (4)	Math (5)	Science (6)	Writing (7)
DPB'_{02}	-0.007 (0.008)	-0.007 (0.005)	4.420 (40.940)	0.013 (0.015)	-0.001 (0.018)	0.013 (0.016)	0.001 (0.016)
Observations	702	696	699	488	305	305	295

Notes: Each outcome is regressed on school-specific DPB'_{02} values as well as indicators for the lottery year. Standard errors (in parentheses) are clustered by application school. Each regression sample is limited to baseline sample restrictions specified in the notes for Table 1.6 and is additionally limited to students who accepted a NCLB position in a 6th grade magnet school from 2002-2007. NCLB rank refers to the district's internal ranking of the lowest-achieving, poorest students in the district. The lower the rank value the higher the priority that student receives for NCLB placement.

Table 1.5: Lottery Estimates of Effects of Magnet Attendance on Teacher, School, and Peer Characteristics

	Teacher/School Characteristics			Student Characteristics of Entire School			
	Fraction Masters (1)	Average Experience (2)	Average Class Size (3)	Fraction Black (4)	Fraction FRL (5)	Math (6)	Reading (7)
Panel A: Pooled Sample							
Enrolled in Magnet (2SLS)	0.020 (0.119) [0.578]	-1.256 (1.000) [11.364]	3.945*** (1.307) [23.035]	0.011 (0.123) [0.808]	-0.202* (0.122) [0.771]	0.434* (0.226) [0.163]	0.284 (0.239) [-0.013]
Lottery Offer (First-Stage)	0.155** (0.062)	0.155** (0.062)	0.155** (0.062)	0.193*** (0.067)	0.193*** (0.067)	0.193*** (0.067)	0.196*** (0.067)
Observations	9,994	9,962	9,970	13,405	13,429	13,403	12,937
Panel B: Subgroup Estimates							
Black	-0.026 (0.150) [0.576]	-0.908 (1.145) [11.311]	4.216*** (1.617) [22.994]	-0.053 (0.129) [0.833]	-0.263* (0.135) [0.777]	0.529** (0.221) [0.153]	0.410* (0.237) [-0.024]
Non-Black	0.153*** (0.057) [0.594]	-1.888** (0.917) [11.740]	3.223*** (1.115) [23.323]	0.198* (0.103) [0.639]	-0.032 (0.078) [0.733]	0.144 (0.199) [0.234]	-0.063 (0.183) [0.060]

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Outcome means among students not offered a magnet lottery seat are in brackets. Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions specified in the notes for Table 1.6. Outcomes in columns 1-3 are averages over the actual classrooms that the given student attended in 6th grade, while remaining outcomes are averages over the entire school. FRL data are from CCD school averages. Regressions are weighted by one over the number of lotteries applied to by the student in the given year.

Table 1.6: Lottery Estimates of Effects of Magnet Enrollment on Student Outcomes

	Achievement Index (1)	Postsecondary Index (2)	Total Index (3)
Panel A: <i>2SLS Estimates for Pooled Sample</i>			
Enrolled in Magnet	0.120 (0.084)	-0.039 (0.128)	0.050 (0.073)
Panel B: <i>2SLS Estimates by Subgroup</i>			
Non-Black	0.147 (0.105)	-0.031 (0.218)	0.037 (0.081)
Black	0.099 (0.091)	-0.060 (0.128)	0.048 (0.079)
Male	0.151 (0.158)	0.102 (0.169)	0.147 (0.136)
Female	0.111** (0.049)	-0.115 (0.164)	-0.005 (0.054)
Above Median Baseline Math Score	0.135 (0.114)	-0.068 (0.178)	0.007 (0.087)
Below Median Baseline Math Score	0.067 (0.140)	0.141 (0.218)	0.123 (0.140)
Black, Male	0.106 (0.168)	0.189 (0.151)	0.182 (0.158)
Black, Female	0.111 (0.068)	-0.202 (0.183)	-0.037 (0.058)
Black, Above Median	0.169 (0.151)	-0.116 (0.195)	0.018 (0.102)
Black, Below Median	0.029 (0.152)	0.150 (0.272)	0.137 (0.166)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions i.e., the student must have applied to a magnet school lottery in the 6th grade from 1998-2007, must be in their first year attending the grade of the lottery application (no grade retention), and must not be eligible for special education. Further, the baseline sample restriction also excludes students from a school with automatic placement into a magnet school. Achievement index is the simple mean of math, science, reading, and writing achievement. Post-secondary index is a simple mean over standardized versions of whether 18 months after high school graduation the student enrolled in any postsecondary institution, a 2-year, 4-year, or Top 50 ranked institution. Regressions are weighted by one over the number of lotteries applied to by the student in the given year. First-stage estimates, observation counts, weak IV tests, and outcome means are provided in Tables C.1, C.2, C.3, and C.4, respectively. See Appendix B for disaggregated estimates.

Table 1.7: Lottery Estimates of Effects of School Peer Racial Composition on Peer Characteristics

	School Peer Composition				
	Baseline Testing				
	FRL (1)	Reading (2)	Math (3)	Science (4)	Writing (5)
Panel A: <i>2SLS Estimates</i>					
Fraction Black in School $\times 100$	0.010*** (0.003)	-0.011** (0.005)	-0.014*** (0.005)	-0.013*** (0.004)	-0.008 (0.005)
Non-offer Outcome Mean	0.695	0.170	0.053	0.050	0.108
Panel B: <i>First-Stage Estimates</i>					
$\mathbb{1}(\text{Initial Offer}) \times \mathbb{1}(\text{Post 2002}) \times \text{DPB}$	2.091*** (0.492)	2.087*** (0.492)	2.187*** (0.496)	2.277*** (0.461)	2.275*** (0.462)
F Statistic (Weak IV)	18.042	17.974	19.424	24.413	24.234
Observations	13,398	13,396	12,930	11,316	11,285

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.4) where each outcome is regressed on the percentage of black students in the grade of the school the student attended during the year following the lottery as well as indicators for student gender, race, year-of-test and lottery. I instrument for endogenous racial composition with a triple-interaction between whether the student received an initial lottery offer, whether the observation is after 2002, and by the 2002 difference in percentage of black students (DPB^{02}) between the total magnet lottery pool and the pool of initial offers. Standard errors are two-way clustered both by student and by school-after-lottery. Because observations are at the student-application level for these regressions, a given student-year combination can appear multiple times. Each regression sample is limited to the sample restrictions specified in the notes for Table 1.6. 2SLS estimates and the outcome mean for students not offered a magnet seat are provided in Panel A. First-stage estimates and weak instrument tests using the Kleibergen-Paap Wald F statistic are reported in Panel B.

Table 1.8: Lottery Estimates of Effects of School Peer Racial Composition on Student Achievement

	Achievement Testing			
	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>2SLS Estimates for Pooled Sample</i>				
Fraction Black in School $\times 100$	-0.007 (0.005)	-0.012*** (0.005)	-0.012*** (0.004)	-0.012*** (0.004)
Panel B: <i>2SLS Estimates by Subgroup</i>				
Black	-0.006 (0.005)	-0.015*** (0.005)	-0.015*** (0.005)	-0.010** (0.004)
Male	-0.007* (0.004)	-0.010** (0.005)	-0.011** (0.006)	-0.012** (0.006)
Female	-0.007 (0.008)	-0.016** (0.007)	-0.012** (0.005)	-0.011** (0.005)
Above Median Baseline Math Score	-0.008** (0.004)	-0.012** (0.006)	-0.013** (0.006)	-0.013 (0.008)
Below Median Baseline Math Score	-0.004 (0.010)	-0.008 (0.008)	-0.011 (0.009)	-0.022* (0.011)
Black, Male	-0.005 (0.004)	-0.014*** (0.004)	-0.016** (0.007)	-0.013** (0.006)
Black, Female	-0.007 (0.009)	-0.017* (0.009)	-0.012** (0.005)	-0.007 (0.006)
Black, Above Median	-0.008* (0.005)	-0.015** (0.007)	-0.020*** (0.007)	-0.010 (0.007)
Black, Below Median	-0.003 (0.010)	-0.007 (0.009)	-0.009 (0.009)	-0.023* (0.013)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.4) where each outcome is regressed on the percentage of black students in the grade of the school the student attended during the year following the lottery as well as indicators for student gender, race, year-of-test and lottery. I instrument for endogenous racial composition with a triple-interaction between whether the student received an initial lottery offer, whether the observation is after 2002, and by the 2002 difference in percentage of black students (DPB^{02}) between the total magnet lottery pool and the pool of initial offers. Standard errors are two-way clustered both by student and by school-after-lottery. Because observations are at the student-application level for these regressions, a given student-year combination can appear multiple times. Each regression sample is limited to the sample restrictions specified in the notes for Table 1.6. First-stage estimates, observation counts, weak IV tests, and outcome means are provided in Appendix Tables C.5, C.6, C.7, and C.8, respectively.

Table 1.9: Lottery Estimates of Effects of School Peer Racial Composition on High School Graduation and Postsecondary Attainment

		College Attendance (18 Months after Graduation)			
	HS Grad. (1)	Any (2)	2-year (3)	4-year (4)	Top 50 (5)
Panel A: <i>2SLS Estimates for Pooled Sample</i>					
Fraction Black in School $\times 100$	-0.004** (0.002)	-0.005** (0.003)	-0.007** (0.003)	-0.000 (0.003)	0.000 (0.000)
Panel B: <i>2SLS Estimates by Subgroup</i>					
Black	-0.004 (0.002)	-0.006* (0.003)	-0.007** (0.003)	-0.000 (0.003)	0.000 (0.000)
Male	-0.001 (0.004)	-0.001 (0.004)	-0.003 (0.003)	0.004 (0.004)	0.001** (0.001)
Female	-0.008** (0.004)	-0.009* (0.005)	-0.011*** (0.004)	-0.002 (0.005)	-0.001** (0.000)
Above Median Baseline Math Score	-0.005 (0.003)	-0.002 (0.002)	-0.007* (0.004)	0.001 (0.003)	-0.000 (0.001)
Below Median Baseline Math Score	-0.006 (0.005)	-0.010* (0.006)	-0.005 (0.005)	-0.001 (0.006)	0.000 (0.001)
Black, Male	-0.001 (0.004)	-0.002 (0.004)	-0.005 (0.003)	0.003 (0.004)	0.001* (0.001)
Black, Female	-0.007* (0.004)	-0.008 (0.006)	-0.012** (0.005)	-0.001 (0.005)	-0.001*** (0.000)
Black, Above Median	-0.006* (0.003)	-0.002 (0.003)	-0.008** (0.004)	0.002 (0.003)	0.000 (0.001)
Black, Below Median	-0.004 (0.004)	-0.014* (0.008)	-0.008 (0.007)	-0.003 (0.007)	0.001 (0.001)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.4) where each outcome is regressed on the percentage of black students in the grade of the school the student attended during the year following the lottery as well as indicators for student gender, race, and lottery. I instrument for endogenous racial composition with a triple-interaction between whether the student received an initial lottery offer, whether the observation is after 2002, and by the 2002 difference in percentage of black students (DPB^{02}) between the total magnet lottery pool and the pool of initial offers. Standard errors are two-way clustered both by student and by school-after-lottery. Because observations are at the student-application level for these regressions, a given student-year combination can appear multiple times. Each regression sample is limited to the sample restrictions specified in the notes for Table 1.6. Top 50 denotes an indicator equal to one if the student attends a top 50 ranked school based on the U.S. News and World Report. First-stage estimates, observation counts, weak IV tests, and outcome means are provided in Appendix Tables C.9, C.10, C.11, and C.12, respectively.

Chapter 2

The Effect of Charter Competition on Unionized District Revenues and Resource Allocation

The charter school movement is rapidly expanding across the United States. Charters are designed to be innovative laboratories for educational practices and to compete with traditional public school districts (TPSD) over student enrollment. Proponents argue that these market forces cause TPSDs to improve student achievement, but the empirical evidence is mixed (Epple et al., 2015). This literature has focused directly on student outcomes instead of the mechanisms underlying how districts respond to charter competition. Without understanding how TPSDs respond, it is difficult to disentangle why competition improves student outcomes in some contexts and reduces outcomes in others. A primary mechanism by which charter competition may operate is through its influence on district resources. Competition may impact the level of overall revenues available to the district and may augment how districts allocate these funds. Moreover, changes to resource allocation decisions provide insight into which dimension of school quality competition affects. For example, districts competing over achievement ratings may allocate resources toward instruction or pupil services, while districts competing over school facility quality may allocate expenditures toward new capital projects. However, we understand little about the extent to which charters

influence TPSD expenditure decisions.¹

Conversely, critics of the charter movement argue that charter competition puts fiscal stress on traditional schools making the remaining students worse off. Empirical evidence confirms that charters place fiscal stress on TPSDs (Bifulco and Reback, 2014) and, in general, decrease the revenues available to districts (Arsen and Ni, 2012b).² Yet, we have an incomplete understanding of why TPSD revenues fall in the presence of charter competition. Some of the decline is mechanical: TPSDs directly lose state per-pupil funding as students transfer to charter schools and federal per-pupil funding as vulnerable student populations transfer. However, other mechanisms may be more nuanced. For example, if charter presence is capitalized into housing values, then charter entry would indirectly affect the TPSD local revenues raised through property taxes.³

This study addresses these gaps in the literature by exploring potential mechanisms underlying how charter competition affects TPSD funding and whether districts respond by adjusting the composition of their expenditures. A core problem this literature has faced is that charter entry is not exogenous with respect to underlying trends in overall TPSD resource levels and allocation decisions. I exploit the fact that Ohio charter entry policies create a substantial lag structure and geographic constraints to isolate plausibly exogenous charter entry variation in a difference-in-difference-instrumental-variables framework. I document that charter competition directly reduces state and federal revenues through the expected channels. A key finding of this study is that charter competition also decreases the TPSD revenues raised through property taxes by depressing appraised district-level residential property values. I also find that charter competition causes districts to spend less on instructional and other current expenditures and spend more on new

¹The only other evidence on within-district resource allocation comes from Arsen and Ni (2012b) who find that charter schools have a negligible effect on TPSD resource reallocation in Michigan. Due to data limitations, Arsen and Ni (2012b) impute charter competition levels for roughly 75 percent of their sample. This can introduce potentially serious attenuation bias into their results and highlights the value of analyzing this question in a setting with a more accurate measure of charter competition.

²Additionally, Dee and Fu (2004) find that charter schools increase pupil-teacher ratios in traditional public schools.

³While Imberman et al. (2016) find no evidence of charter capitalization on average in Los Angeles county, they find that housing prices outside the Los Angeles Unified School District fall in response to within-district charter entry.

construction capital outlays. This reallocation is more than a simple proportional change. A one percentage point increase in charter competition increases the overall amount that TPSDs spend on capital outlays by 7.3 percent. This is consistent with qualitative evidence that administrators in Washington D.C. believe the physical appearance of their school has the greatest impact on preventing enrollment loss to charters (Sullivan et al., 2008). I discuss further explanations for these surprising results below. Additionally, I provide evidence that these findings are not driven by the passage of the No Child Left Behind Act nor the Great Recession.

I also examine the effect of charter competition on collectively bargained teacher salaries. Most studies of the effect of charter competition on teacher salaries occur in settings where collective bargaining is prohibited by law, such as Texas (Taylor, 2006, 2010) and North Carolina (Jackson, 2012).⁴ Thus, I address a gap in the teacher labor market literature by assessing how unionized markets respond to largely non-unionized charter school competition. A challenge in studying the effect of charter competition on collectively bargained teacher salaries is that contracts are negotiated intermittently and can only adjust to charter competition during negotiation years. Ignoring this problem generates an attenuation bias.⁵ The bias is similar to the well-known “seam bias” in popular panel datasets such as the Survey of Income and Program Participation, which arises when respondents answer retrospective questions using current information (see Ham et al., 2009; Pei, 2015; Pischke, 1995).

I characterize this bias within my context and use Monte Carlo simulations to demonstrate that the bias is avoided by restricting the analysis sample to years when the outcome can vary (i.e., negotiation years). Using this approach and the universe of Ohio teachers’ union contracts, I estimate that a percentage point increase in charter competition decreases teacher salary contracts at the top of the pay scale by around 1.0 percent. I also estimate salary decreases for entry-level teachers but find that charter competition has no effect on mid-career salary contracts. Furthermore, I find that charter competition causes TPSDs to hire fewer new teachers, which reduces the size of the teacher labor force to maintain pupil-teacher ratios. I estimate minor teacher mobility between TPSDs and charters consistent with a model where TPSDs are competing over students instead of

⁴Notable exceptions include Arsen and Ni (2012b) and Hoxby (2002).

⁵In my setting, estimates on annual data would theoretically attenuate my results by 7 percent.

teachers.

The remainder of the paper is organized as follows. Section 2.1 overviews Ohio charter school institutional details, Section 3.4 describes the data, Section 2.3 presents the research design and evaluates its validity, and Sections 2.4 through 2.7 provide the main results for district revenues and expenditure allocation emphasizing collectively bargained teacher salaries. Section 2.8 discusses these results and Section 3.8 concludes.

2.1 Institutional Details

Charter schools are independently run educational organizations that sign a “charter” declaring their structure and outlining detailed plans for achieving student success. Charter schools in Ohio differ from traditional public schools in the following ways. While students may only attend a traditional public school based on the geographic location of their residence, students across the state are able to attend any charter they desire.⁶ When a student transfers to a charter from a public school their per-pupil state funding transfers as well. Any charter failing to attract the number of students needed to at least fund operating costs will eventually close.

In Ohio, there is an important distinction between a *conversion* and *start-up* charter school (ODE, 2014). The conversion schools are created by “converting” all or a portion of an existing public school into a charter school. These schools must obtain a majority vote at the school board to convert. Public schools can convert at any time across the state, conditional on receiving the necessary votes. These districts operate independently from their sponsor school district. Conversion charter schools are free to decide if they want to remain unionized.

Start-up schools on the other hand are new educational institutions and differ from conversion schools in a variety of ways. First, start-up charters can be sponsored by a larger set of entities. Sponsors for start-ups can include teachers, parents, communities, private organizations, Ohio universities, and even the Ohio Department of Education (ODE). Start-ups must privately fund

⁶Local school districts are required to provide transportation to any student living more than two miles away from their desired charter school as long as the charter is no further than 30 minutes away from the school of residence.

a majority of the charter’s expenses including the large entry costs. As a result, they often try to renovate and locate in closed-down schools or shopping centers (Imberman, 2011) instead of constructing new buildings. Unlike conversion schools, start-ups are not able to open freely across the state. There is a complicated legislative history (see Section 2.3.2) that dictates in which districts start-up charters are permitted to open during any given year.

Ohio charter schools can be further categorized as either a traditional “brick-and-mortar” or a “digital” charter. Digital charter schools face the same legislation and requirements as traditional “brick-and-mortar” charters; however, all instruction occurs online, and schools are required to provide each student with a laptop. Ohio has the second-largest (second to Arizona) online charter presence in the nation, with over 30,000 students enrolled in a digital school in 2011-12. This represents rapid growth considering the first digital charter school opened in the 2000-01 school year. While there are a handful of digital charter schools that limit enrollment to district residents only, nearly every digital charter allows students from across the state to enroll.

2.2 Data

2.2.1 Data Description

To test the effect of charter competition on district revenues, resource allocation, negotiated contract outcomes, and teacher employment requires several datasets, each of which are summarized below in turn. Additional information is presented in Online Appendix A, which summarizes the details of each dataset including important data cleaning procedures.

Digitized union contracts are provided by the Ohio State Employment Relations Board (SERB). I observe all contracts from 1982 through the 2012-13 school year. About 95 percent of TPSDs first began collective bargaining negotiations between 1984 and 1987. The unit of observation in these data is a district, contract, salary-track observation, where salary tracks are broken out by teacher education (e.g., bachelor’s or master’s degree). To make this explicit, I present a fictitious contract in Online Appendix Table A.2. A teacher’s pay is determined solely based on their years of experience and education level. Notice that payment increases may not necessarily occur each

year, for example, consider the payments for years of experience 14-15. For each district’s contract, I observe the entry- and top-level salaries (those corresponding with experience rows 0 and 28 in Table A.2). However, for most contracts, SERB data custodians instead recorded the top-level salaries as the first year in which a salary does not change with experience, introducing additional noise into this measure. In Table A.2, this corresponds to the bold values, i.e., 8 years of experience for “Non-degree” teachers and 14 years of experience for all other education categories. Beyond salary information, I observe contract information concerning the negotiated number of hours in the work week, and the number of steps and years to reach the top of the pay scale.

In order to fill in information for mid-range and true top-level negotiated salaries, I turn to restricted-access, teacher-level data provided by the Ohio Department of Education (ODE). These data include teacher salary, experience, education, and current school of employment. Importantly, these data follow teachers as they move between public schools within the state, including charters. Furthermore, I observe all necessary information (i.e., teacher experience and education) to determine each teacher’s particular position on their district’s pay scale. As a result, I can use SERB contract negotiation dates in tandem with these teacher-level data to back out entire salary structures for each negotiated contract. This allows me to estimate the effect of charter competition across the entire negotiated salary distribution (see Online Appendix B).

To measure charter competition, I collect public school finance reports for the universe of Ohio school districts from the ODE’s school finance website. In Ohio, when a student decides to attend a charter school instead of his or her default public school, the district must directly pay the baseline per-pupil state funding amount to the charter. The ODE finance reports capture these exact payments as well as a full-time-equivalency count of the number of students each district sends to each charter school in the state, including digital charters.

In addition, I employ information from three National Center for Education Statistics data sources: the School District Universe Survey, School Building Universe Survey, and School District Finance Survey. The first two datasets provide information about student enrollment and teacher employment (see Online Appendix Table A.1 for specifics). The School District Finance Survey provides TPSD revenue and expenditure information.

Finally, I utilize property tax data from the Ohio Department of Taxation, which includes annual property valuations broken out at the district level. These valuations determine the property tax base for TPSD local revenue calculations.

2.2.2 Measuring Charter Competition

I measure charter competition using information about the number of students transferring from TPSDs to charter schools (Linick, 2014).⁷ Specifically, my preferred measure of charter competition is the fraction of a district’s potential membership that instead transfer to a charter school in the given year, i.e., $\frac{\# \text{ Transfers to Charter}}{\# \text{ Students Enrolled in TPSD} + \# \text{ Transfers to Charter}}$.⁸ This measure of competition includes student transfers to conversion, start-up, digital, and brick-and-mortar charters (see Section 2.1).

In 2001, the ODE began generating “District Foundation Settlement Reports” that record the number of students and accompanying funds sent by each district to each charter school across the state, including digital charters. For 2001 and later, I use this full-time equivalency count as my measure of the number of students transferring to charters from each district.

From 1998 to 2001, in order to estimate a proxy for charter transfers, I use information from several sources. Competition is proxied using the amount of district-aggregated funds transferred to charter schools each year as recorded in the Common Core of Data (CCD) School District Finance Survey. To convert these dollar values into counts of transferring students, I collected information on the baseline formula dollar amounts paid to a charter school for transferring a single student each year. Note that unlike the post-2001 data, these measures only provide the general competition a specific district faces making it impossible to disaggregate transfers between digital and brick-and-mortar charters. For twenty-one district-year observations with missing CCD charter payment information prior to 2001, I fill in the district’s charter transfers with the cumulative enrollment counts for all charters “serving” the given district.⁹

⁷Notice that student-transfer measures do not capture charter threat. For example, a charter opening within TPSD boundaries that was unable to recruit students from the given TPSD would be measured as contributing no competitive pressure. In Online Appendix C, I show that my main estimates are robust across a variety of charter competition measures.

⁸Potential enrollment and charter transfers are in full-time equivalency terms.

⁹While students are free to attend any charter across the state, charters are tied to a “serving” district. The

In Figure 2.1, I plot the aggregated enrollment counts from the CCD LEA Universe Survey for all charter schools in the state, as well as an aggregated version of my measure of charter transfers. From 2001 onward, even though my measure of charter competition is not based on CCD data, I am able to closely mirror aggregated CCD charter enrollment. I take this as evidence that my measure of charter competition adequately captures statewide charter enrollment. For 2001 and later, Figure 2.1 also decomposes the number of charter transfers into the number transferring to brick-and-mortar and digital charter schools. In 2001, charter transfers were almost entirely made up of brick-and-mortar schools, with the share of the charter market captured by digital charters steadily growing over time.

The spatial growth as measured by the fraction of TPSD enrollment that instead transfers to a charter school is shown in Figure 2.2. Because the most urban districts have always been eligible for charter entry, charter school hot-spots appear in Ohio’s largest cities. However, legislation passed in 1999 and 2002 that allowed charters to open in struggling districts across the state coupled with statewide digital charter admissions generate large increases in the fraction of charter transfers across rural Ohio in later years.

Taken together, Figures 2.1 and 2.2 show why Ohio is an excellent state to assess the effects of charter competition. There has been rapid and relatively recent charter introduction and expansion, which provides the necessary treatment variation. These facts coupled with the policy structure that generates exogenous variation in the timing and location of charter entry make Ohio an ideal setting to study the competitive effects of charter schools.

Table 2.1 presents descriptive statistics for TPSDs broken out by various levels of charter competition intensity. The top panel reports district characteristics and finance information for outcomes that vary at the district-year level. Column 1 provides the mean and standard deviation for the full sample of district-year observations that have non-missing values for all variables in the panel. Columns 2 through 4 present summary statistics for district-year combinations facing no charter competition, non-zero levels of competition, and the top quartile of charter competition, respec-

eligibility of this district originally determines whether the charter could open. This assumption is reasonable particularly for the earlier years of charter entry. In 2001 and 2002, over three-fourths of all districts sent 65 percent or more of their charter transfers to charters specifically serving their district.

tively.

Comparing across columns, districts facing higher levels of competition are larger (enroll more students and employ more teachers), have higher assessed total property values, have a higher proportion of black students as well as students qualified for free lunch, spend more overall, and have higher instructional and capital expenditures. I also report descriptive statistics for the amount of charter competition faced by each TPSD. It is worth noting that for the full sample, on average, the fraction of potential enrollment transferring to charters is roughly 0.014. This fraction more than triples for districts facing the top quartile of competition. In the bottom panel, I report entry- and top-level salaries negotiated between districts and unions. The unit of observation in this panel is a contract-education cell. The pattern is not as clear across these variables, but districts facing competition tend to negotiate higher salaries. Overall, the main empirical results of this paper are not visible in these simple tabulations.

2.3 Methodology

2.3.1 Baseline Estimates: Difference-in-Difference Framework

To understand the intuition behind my baseline empirical setup, consider two districts within the same local economy in a given year where one experiences increased charter competition in the following year and the other does not. My baseline estimation strategy simply compares the change in outcomes over time between these districts. Specifically, I estimate the following model:

$$y_{ict} = \alpha + \beta C_{it} + \gamma_{ct} + \phi_i + \epsilon_{ict} \quad (2.1)$$

where y_{ict} is the outcome of interest for district i during school year t in commuting zone c , γ_{ct} are commuting-zone-by-school-year fixed effects, ϕ_i are district fixed effects, and ϵ_{ict} is an idiosyncratic error term.¹⁰ C_{it} denotes the charter competition faced by district i during the school year t .

¹⁰Year 2000 Commuting Zones are downloaded from the Department of Agriculture ([website](#)), which are designed to delineate the local economies where people work and live.

Standard errors are clustered on districts. This setting accounts for year-specific shocks affecting all districts across a given commuting zone as well as time-invariant district characteristics. The identifying assumption of equation (2.1) is that conditional on the fixed effects, charter competition is uncorrelated with any other determinants of the outcome.

Because students choose which school to attend as well as the timing of transfers, any trends in factors driving these choices that also correlate with trends in district outcomes are sources of bias. For instance, suppose that charters tend to locate near districts with downward trending student performance. Further, suppose that these districts experience state sanctions that restrict their budgets and induce changes to resource allocation. Without accounting for district performance trends, charter entry into these downward-trending districts would correlate spuriously with changes in district fiscal outcomes. The commuting-zone-by-school-year fixed effects γ_{ct} help account for unobservable factors potentially correlating with trends in outcomes and charter entry by forcing comparisons only to be made between districts within the same commuting zone.¹¹ In the previous example, shifts in quality that affect all districts within a commuting zone are absorbed by γ_{ct} , but differential quality shifts within a commuting zone would still potentially induce biases.

2.3.2 Preferred Estimates: Instrumental Variables Framework

To account for potential unobservable trends that would bias my baseline estimates, I exploit both the lengthy charter approval process as well as plausibly exogenous changes to policies that determine the location and timing of charter entry in an instrumental variables framework.

In 1997, Ohio legislators passed a bill that, in addition to piloting a new start-up charter program in Lucas county, allowed new start-ups to open in the “Big 8” urban districts (Ohio HB 55).¹² This bill also allowed conversion charter schools the option to open across the state. In 1999, another bill (Ohio HB 282) passed that allowed start-up charters to open in the twenty-one largest

¹¹A specification only including year fixed effects would implicitly be assuming that charter transfer intensity varies exogenously across the entire state. Including commuting-zone-by-school-year fixed effects instead relies on the assumption that charter transfer intensity varies exogenously across school districts within commuting zones.

¹²The “Big 8” urban districts are comprised of Akron, Canton, Cincinnati, Cleveland, Columbus, Dayton, Toledo, and Youngstown.

urban districts. Further, starting in the 2000 school year, start-ups across the state could open in any district rated as “Academic Emergency” (AE) in the previous school year based on Ohio’s performance index rating system.¹³ In 2003, legislation passed that allowed start-up charters to open in any districts rated as “Academic Watch” (AW) or AE in the previous school year, but the bill again limited new start-up charters to open in the “Big 8” districts (down from 21 eligible districts) without regard to the previous year’s performance rating (Ohio HB 364 and HB 3). These designations only affect whether charters are permitted to enter a particular district. Once opened, charters are allowed to persist without regard to their district’s current eligibility status.

Table 2.2 provides the number of districts eligible for new charter entry in the given year based on “Urban 8/21” policies in column 1 and district ratings during the previous school year in column 2.¹⁴ Column 3 presents the total number of districts eligible for new charters to begin the approval process to eventually open within the TPSD. It is worth noting that most of my identifying variation is coming from the 2000-2005 school years, which corresponds to the years with the largest amount of charter growth (see Figure 2.1). Column 4 presents the number of new charters that actually open during the subsequent school year based on district eligibility during the given year. As expected, the number of new charters tracks closely with the number of eligible TPSDs. For example, 2003 was the first year that districts with a lagged AW designation were eligible for charter entry, which led to the sharp increase in number of eligible districts. Based on the eligibility of these 78 districts in 2003, 90 new charters opened up in 2004, representing roughly a three-fold increase over the previous year.

In order for a start-up charter school to open, there is a very specific timeline that must be followed. This timeline is graphically depicted in Figure 2.3a for district-year outcomes. In general, Ohio policies create a one-year lag from when a district is eligible for charter entry to when the new charter can open its doors.¹⁵ Thus, denoting time with respect to an outcome in school year t ,

¹³Performance Index ratings are calculated by taking a weighted average of the fraction of students who passed different statewide goals. See Online Appendix D for a detailed explanation of the ratings designation system in Ohio.

¹⁴“Urban 8/21” districts that are also rated as AE/AW only appear in column 1.

¹⁵Charter schools are required to first enter into a Preliminary Agreement with an authorizer before proceeding to finalize a charter or contract. When a charter enters into a Preliminary Agreement, they must identify their intended district of residence, which must be eligible for charter entry at the time the Preliminary Agreement is

a new charter opening in period t would have needed to initiate the filing process in the previous school year (denoted $t - 1$). Further, district eligibility in $t - 1$ depends on whether it is a “Big 8”/“Urban 21” district in $t - 1$ or based on the district’s academic rating in the previous school year, $t - 2$.

I use as instruments the change in the differential effects of being rated in AE(AW) before and after the introduction of the 2000(2003) policies that introduced using district rating criteria to determine charter eligibility. I operationalize this by interacting both $t - 2$ lagged district rating indicator variables with an indicator for whether the given policy had been implemented prior to or concurrent with the $t - 2$ school year. I then use both interactions as instruments for charter entry and include the main effects as controls. As my third instrument, I include a binary for whether the district was eligible for charter entry during the previous year based on urban district policies (“Big 8/Urban 21”).

While the timing of the urban district eligibility policies are arguably exogenous, it is plausible that districts receiving an AE/AW rating during the previous year could be subject to correctional responses that possibly correlate with district resource allocation and negotiated salaries (see [Craig et al., 2013](#); [Chakrabarti, 2014](#)). Suppose that outcomes react directly to $t - 2$ AE/AW ratings. In my empirical framework, I control for this directly and assume that there is no other structural break besides from the policy introduction. The identifying assumption underlying this strategy is that the only thing changing the relationship between $t - 2$ AE/AW ratings and subsequent outcomes is the passage of the relevant charter law.

executed. Because district ratings are released September 15th each year, TPSD eligibility is based off ratings during the previous school year. Even if the district’s status changes with the release of the next state Report Card, the new charter school is still permitted to open. The Preliminary Agreement must be signed by the end of the December preceding the school year in which the new charter plans to open. Typically, Preliminary Agreements do not extend beyond twelve months. However, the contract/charter has a very specific time frame – it must be adopted no later than March 15th prior to the school year in which the school intends to open, and must be executed by May 15th following the March adoption date.

The school must open by September 30th following the contract execution date, unless it is a school serving primarily dropout prevention and recovery students, in which case the contract is valid for 12 months from the execution date. Generally the whole process takes less than 12 months.

In addition, I exploit the delay in the charter approval process. Suppose that poor district performance affects outcomes through two channels: first, through its effect on charter entry several years into the future, and second, through its more immediate, direct effect on district outcomes. The first channel provides useful identifying variation. To isolate this variation, I include two binary controls, each respectively equal to one if the given district receives an AE or AW rating during the previous period, $t - 1$.¹⁶ In this setting, the more immediate, direct effects of poor district performance are absorbed by the $t - 1$ district rating indicators and the IV is identified off of the lagged structure of charter entry.

Specifically, I regress:

$$y_{ict} = \beta C_{ict} + \delta_1 \mathbb{1}(AW)_{i,t-1} + \delta_2 \mathbb{1}(AE)_{i,t-1} + \phi_1 \mathbb{1}(AW)_{i,t-2} + \phi_2 \mathbb{1}(AE)_{i,t-2} + \gamma_{ct} + \eta_i + \epsilon_{ict} \quad (2.2)$$

using the corresponding first stage

$$\begin{aligned} C_{ict} = & \kappa_1 \mathbb{1}(AW)_{i,t-1} + \kappa_2 \mathbb{1}(AE)_{i,t-1} + \xi_1 \mathbb{1}(AW)_{i,t-2} + \xi_2 \mathbb{1}(AE)_{i,t-2} \\ & + \theta_1 \underbrace{\mathbb{1}(AE)_{i,t-2} \cdot \mathbb{1}(\text{Post } 1999)_{t-2}}_{\text{HB 282}} + \theta_2 \underbrace{\mathbb{1}(AW)_{i,t-2} \cdot \mathbb{1}(\text{Post } 2002)_{t-2}}_{\text{HB 364}} \\ & + \theta_3 \underbrace{\mathbb{1}(U)_{i,t-1}}_{\text{HB 3, 55, 282}} + \Gamma_{ct} + \psi_i + \nu_{ict} , \end{aligned} \quad (2.3)$$

where y is an outcome for district i in school year t and commuting zone c . C is a measure of charter competition faced by district i in year t .¹⁷ $\mathbb{1}(AW)$ is a binary equal to one if the district was in “Academic Watch” one period ago, denoted with subscript $t - 1$ or two periods ago, denoted $t - 2$. $\mathbb{1}(AE)$ is similarly defined, but for districts previously rated as being in “Academic Emergency”. $\mathbb{1}(\text{Post } 1999/2002)$ denote binaries equal to one if the $t - 2$ school year occurred on or after 1999

¹⁶Recall that charter potential during period $t - 1$ is a function of $t - 2$ district ratings, allowing me to separately control for period $t - 1$ academic ratings.

¹⁷Charter competition is measured with error. To the extent that the instruments are correlated with the true level of charter competition and are uncorrelated with measurement error, estimating (2.2) will correct for the mismeasurement.

and 2002, respectively.¹⁸ $\mathbb{1}(U)$ denotes a binary for whether the district qualified for new charter entry in the previous year by being one of the urban eight/twenty-one districts or a district in Lucas county. All other variables are as previously defined.

This setup is embedded within the baseline difference-in-difference framework from (2.1). Thus, the remaining threat to validity comes from any change in the relationship between lagged AE(AW) status and outcomes before and after 1999(2002) that is not due to the introduction of the given charter policy across districts within a given commute-zone.

No Child Left Behind and the Great Recession

In 2002, the No Child Left Behind Act (NCLB) was signed into law as an update to the Elementary and Secondary Education Act of 1965. Because the introduction of NCLB accountability measures overlap with the implementation of the aforementioned Ohio charter policies, NCLB presents a potential concern for the validity of my identification strategy. Under NCLB, districts were rated by whether they met indicators based on the percent of various student subgroups passing standardized tests. Districts were considered to be on an acceptable trajectory if they met their “Adequate Yearly Progress” (AYP) requirements and schools/districts consecutively failing to meet AYP requirements received increasingly harsh federal sanctions as failing tenure increased.¹⁹

AYP criteria play only a minor roll in the determination of AE/AW ratings in Ohio and thus have limited potential to affect TPSD charter eligibility.²⁰ However, because NCLB was implemented within a similar window to the charter policies I exploit, it is plausible that the introduction of NCLB could directly affect how TPSDs allocate their resources outside of its effect on charter school

¹⁸The 2000 law used AE status as a criterion for TPSD charter eligibility starting with the 1999 school year. The 2003 policy made TPSD charter eligibility reliant on AW status starting with the 2002 school year. Hence, I interact $t - 2$ TPSD ratings with $t - 2$ 1999 and 2002 indicators.

¹⁹Sanctions included setting aside part of a school’s Title I funding to allow students to transfer out of their school of residence and to provide free tutoring. Schools consecutively failing AYP even risked closure.

²⁰Specifically, districts with Ohio-specific state indicators between 50-74.9 percent or that have a performance index score of 80-89.9 are rated as “Continuous Improvement” if they meet AYP or “Academic Watch” if they fail AYP. For all other cases, AYP status cannot change a district’s final categorical rating (See Online Appendix Figure D.1).

transfers creating a potential bias in my instruments. With that said, my $t - 1$ lagged district rating information will partially control for any NCLB direct effects. Further, while NCLB was signed into law in 2002, many of the sanctions could not be implemented until schools/districts repeatedly failed AYP, potentially making the full-impact of NCLB not felt until around 2004 to 2005.

In order to test the sensitivity of my estimates to potential NCLB contamination, in Online Appendix E, I present two sets of robustness checks. The first set attempts to control directly for NCLB policies using the entire regression sample, while the second uses my original specifications but limits the sample to pre-NCLB school years. Incidentally, because the Great Recession occurred several years after the passage of NCLB, this set of specifications also tests the extent to which the Great Recession may be driving my results. Estimates are stable across both sets of robustness checks providing evidence that NCLB sanctions and the Great Recession do not present a first-order validity concern.

Interpreting the Local Average Treatment Effect

Due to the complicated instrument structure, it is worth carefully describing the subset of districts that identify the local average treatment effect (LATE) (Imbens and Angrist, 1994). My empirical strategy estimates a LATE from the population of schools that were eligible in the previous period for new charter entry based on either low academic ratings or large urban district categorical eligibility. As a result, my estimates provide the causal effect of charter competition specifically for these low-performing districts and will miss any heterogeneous charter effects at different points of the district performance distribution.²¹

Further, I am only identifying the effect of charter entry from charters serving a given TPSD. Specifically, the instruments leverage only increased charter transfers resulting from the charter potential status of the given district. To see this, consider a TPSD that was ineligible for new charter entry last period but is neighboring a district that recently opened a new charter school. Even if the TPSD sends students to the new charter in the neighboring district, because the TPSD

²¹Specifically, the LATE is an efficiently weighted average of the causal effects for districts made eligible from either low academic ratings or urban district categorical eligibility (Angrist et al., 2016). However, because urban districts often receive poor academic ratings, my estimates will primarily reflect effects for low-performing districts.

was ineligible for new charter entry these transfers do not contribute identifying variation. Thus, my strategy also abstracts from estimating the effect of any inter-district spill-overs of charter competition.

2.3.3 Adjusting Methodology for Contract Outcomes

Union contract outcomes have a fundamentally different data structure than the district-by-school-year outcomes discussed in 2.3.2. As a result, to estimate the effect of charter competition on contract outcomes, I must augment the estimation procedure. One important difference between contract outcomes and district-by-school-year outcomes is that union contracts are negotiated intermittently instead of annually. Below, I describe the bias that would result had I attempted to assess annual teacher salary measures from staff data instead of intermittent contract measures. Following this discussion, I detail how I adjust my estimation framework for collectively bargained contract outcomes.

Mechanical Bias from Partially Fixed Outcomes

In this section, I demonstrate that even ignoring the biases arising from selection concerns, a standard approach to estimating the effect of competition on negotiated salaries is mechanically biased if the researcher treats salaries as though they vary annually when in reality the contracts are negotiated intermittently.²² In Online Appendix F, I derive a closed-form solution for the amount of mechanical bias that results from treating the dependent variable as though it can vary in each period, when in reality, it only varies periodically.

Specifically, if I denote β as the true parameter value for the variable of interest, x , with the accompanying estimate $\hat{\beta}$, the mechanical bias is given by

$$\hat{\beta} = \beta[1 - \underbrace{\delta(1 - \rho_x(g))}_{\text{Mechanical Bias}}] \quad (2.4)$$

²²For example, [Vedder and Hall \(2000\)](#) study the effect of private school competition on teacher salaries in Ohio and treat salaries as if they vary annually.

where δ is the fraction of outcome observations fixed in the sample, but treated as if they vary and $\rho_x(g)$ is the correlation coefficient between x and g lags of x . Because $\rho_x(g) \in [-1, 1]$ the estimate can either overstate or understate the truth depending on the autocorrelation in x . However, in applications using data with a positive serial correlation in x , the bias attenuates estimates. Notice, that the bias disappears if $\delta = 0$, i.e., that all outcomes vary annually, or if $\rho_x(g) = 1$, i.e., x is perfectly serially correlated so that x values during a year when the outcome can vary are a perfect representation of the x values during the fixed outcome years. In Online Appendix F.1, I provide evidence from Monte Carlo simulations that the predicted bias from (2.4) matches the estimated bias when the truth is known. I also show that restricting the sample to only observations in which the outcome can vary completely mitigates the bias.²³

This bias is similar to the well-known “seam bias” arising from telescoping behavior of respondents in important retrospective panels such as the Survey of Income and Program Participation (see Pischke, 1995; Ham et al., 2009; Pei, 2015).²⁴ While not directly applicable to duration or event study models as estimated in Ham et al. (2009) and Pei (2015), any researcher interested in estimating the effect of some x on y using retrospective panels can potentially mitigate the bias resulting from telescoping behavior by restricting the sample only to observations from the month the survey was collected and omit observations from retrospective months. However, further work is needed to formalize this extension and fully characterize its implications for retrospective panels.

In the setting of this study, contracts are negotiated roughly every three years (i.e., $\delta \approx 0.66\bar{7}$) and the serial correlation in charter competition between non-contract years and the corresponding previous negotiation year is roughly 0.9 (i.e., $\rho_x \approx 0.9$). Thus, estimates of the effect of charter competition on annual measures are predicted to be attenuated by about 7 percent. In order to avoid this mechanical bias when estimating the effect of charter competition on collectively bargained wage contracts, the researcher must observe the contract negotiation dates to correctly specify the years in which observed outcomes are able to adjust. As a result, my empirical analysis for contract outcomes will be conducted on only the years with newly negotiated contracts.

²³For example, Card (1990) studies the effect of previous collectively bargained wage rates on subsequent wages by assessing union contract outcomes directly.

²⁴Telescoping occurs when respondents answer retrospective questions using information from the present.

Difference-in-Difference Framework

A contract-start-year-by-district uniquely identifies a particular contract. Further, there is a separate salary structure for teachers with different levels of education.²⁵ For brevity, I designate each of these categories as different “salary tracks.” Thus, the unit of observation for a given pay scale step (e.g., entry or top salary) is at the district-by-salary-track-by-contract-start-year level and regressions are limited only to the start years of new contracts. For my baseline specification, I estimate a model similar to equation (2.1). Specifically, I estimate

$$y_{isc\tau} = \alpha + \beta \cdot C_{i,\tau-1} + \gamma_{c\tau} + \xi_s + \phi_i + \epsilon_{isc\tau} \quad (2.5)$$

where $y_{isc\tau}$ is a contract outcome variable occurring during the *contract-start-year* τ for district i , commuting zone c , and salary track s , $\gamma_{c\tau}$ are contract-start-year-by-commuting-zone c fixed effects, ϕ_i and ξ_s are respectively district and salary-track fixed effects, and $\epsilon_{isc\tau}$ is the error term. $C_{i,\tau-1}$ denotes the charter competition faced by district i during the school year prior to the contract start year τ (denoted $\tau - 1$). The construction of this variable accounts for the fact that contracts are negotiated during the year prior to the contract start year. I choose to model charter competition as the amount that would be faced during the year that the contract is negotiated as opposed to the year the contract is enforced. Standard errors are clustered by district.

Instrumental Variables Framework

For contract outcomes, I adjust (2.2) by adding salary track fixed effects ξ_s . Further, I add an additional lag for all right-hand-side variables because the relevant charter competition is now in the school year prior to the contract-start-year $\tau - 1$ (see Figure 2.3b). Specifically, I regress:

$$\begin{aligned} y_{isc\tau} = & \beta C_{isc,\tau-1} + \delta_1 \mathbb{1}(AW)_{i,\tau-2} + \delta_2 \mathbb{1}(AE)_{i,\tau-2} \\ & + \phi_1 \mathbb{1}(AW)_{i,\tau-3} + \phi_2 \mathbb{1}(AE)_{i,\tau-3} + \gamma_{c\tau} + \xi_s + \eta_i + \epsilon_{isc\tau} \end{aligned} \quad (2.6)$$

²⁵There is a separate salary schedule for teachers with no degree, a Bachelor’s Degree, a Bachelor’s Degree and 150 semester hours, a Master’s Degree, a Master’s Degree and 15 additional graduate semester hours, a Master’s Degree and 30 additional graduate semester hours, and a Doctoral Degree.

The first stage is given by

$$\begin{aligned}
C_{isc,\tau-1} = & \kappa_1 \mathbb{1}(AW)_{i,\tau-2} + \kappa_2 \mathbb{1}(AE)_{i,\tau-2} + \xi_1 \mathbb{1}(AW)_{i,\tau-3} + \xi_2 \mathbb{1}(AE)_{i,\tau-3} \\
& + \theta_1 \mathbb{1}(AE)_{i,\tau-3} \cdot \mathbb{1}(\text{Post } 1999)_{\tau-3} + \theta_2 \mathbb{1}(AW)_{i,\tau-3} \cdot \mathbb{1}(\text{Post } 2002)_{\tau-3} \\
& + \theta_3 \mathbb{1}(U)_{i,\tau-2} + \Gamma_{c\tau} + \varphi_s + \psi_i + \nu_{isc,\tau-1} \quad .
\end{aligned} \tag{2.7}$$

2.3.4 Validating the Instrumental Variables Strategy

Because my instruments largely exploit variation in districts at the bottom of the performance distribution, one concern about my identification strategy is that these districts are trending downward for reasons outside of charter competition and that these unobservables correlate both with trends in charter transfers and district finance.

I validate my empirical strategy by comparing the effects of a district receiving an “Academic Watch” (AW) rating before and after the introduction of the 2003 charter entry policy for a variety of outcomes. Recall the 2003 policy authorized charters to open within districts receiving an AW rating during the previous school year. If my empirical strategy successfully isolates the variation in charter entry driven by lagged AW status, then the effect of lagged AW ratings on any outcome relative to 2003 should be zero for 2002 and earlier and then potentially non-zero thereafter. I focus on the 2003 policy because I have the longest panel of pre- and post-policy data. Ohio first began implementing this particular rating system in 1998, making 2000 the first year I am able to observe a two-period lag of AW ratings. I implement this test by regressing

$$y_{ict} = \sum_{\substack{t=2000; \\ t \neq 2003}}^{2011} \{ \beta_t \mathbb{1}(AW)_{i,t-2} \times \mathbb{1}(\text{Year} = t) \} + \delta \mathbb{1}(AW)_{i,t-2} + \gamma_{ct} + \eta_i + \varepsilon_{ict} \quad , \tag{2.8}$$

where y_{ict} is the given outcome for district i during school year t , in commute zone c , and $\mathbb{1}(AW)$ and fixed effects are defined as in (2.2). Fixed effects are included to net out unobservables already accounted for by the baseline difference-in-difference specification.²⁶ Each β_t captures the year-specific effect of receiving an AW rating two years prior to school year t . In 2003, TPSDs were

²⁶The year main effects are absorbed by γ_{ct} .

first eligible for charters to begin the paperwork to enter based on AW status. Thus, new charters entering from the 2003 law would be able to open as early as the 2004 school year. As a result, I benchmark all estimates relative to 2003. For contract outcomes, I adapt equation (2.8) using the notation from equation (2.6) to estimate

$$y_{iscr} = \sum_{\substack{t=2001; \\ t \neq 2004}}^{2011} \{ \beta_{\tau} \mathbb{1}(AW)_{i,\tau-3} \times \mathbb{1}(\text{Contract-start-year} = \tau) \} + \delta \mathbb{1}(AW)_{i,\tau-3} + \gamma_{c\tau} + \eta_i + \xi_s + \varepsilon_{iscr} . \quad (2.9)$$

Notice that these validity checks only exploit a portion of my identifying variation. The full specification benefits from the additional power provided by the urban district and 2000 AE eligibility policies.

Figure 2.4a displays the time-varying effects of receiving an AW rating two years earlier on the fraction of students transferring out of the district. The difference between the average effects before and after 2003 provide a visual for the first-stage estimates of the policy. For earlier school years, the differential effects of a lagged AW rating on charter entry is statistically indistinguishable from the 2003 effect. For 2004 and later, a lagged AW rating generates significantly more charter entry than in 2003. This pattern supports the validity of my instrument because charter entry is only driven by lagged AW ratings after the relevant policy change.

In the subsequent panels of Figure 2.4, I present the differential effects of lagged AW status on various outcomes, which depict the reduced-form estimates of the AW policy. While individual estimates are often underpowered, the overall trends are still illuminating.²⁷ First, in Figure 2.4b, I plot the effect of a district receiving an AW rating on the inverse hyperbolic sine (IHS) of district-level appraised property values.²⁸ This measure could reflect information about how the economy as a whole is affected over time. For the pre-2003 years, the effect of being rated in AW two years

²⁷For the later years in the sample, few districts were eligible for charter entry based on lagged AW ratings (see Table 2.2).

²⁸The inverse hyperbolic sine of y is $\sinh^{-1}(y) = \ln(y + \sqrt{y^2 + 1})$. Note that $\sinh^{-1}(0) = 0$ and similar to the natural log transformation, $\frac{\partial \sinh^{-1}(y)}{\partial x} = \left(\frac{1}{\sqrt{y^2 + 1}} \right) \frac{\partial y}{\partial x} \rightarrow \frac{1}{y} \left(\frac{\partial y}{\partial x} \right)$ as $y \rightarrow \infty$ (see footnote 17 of Cascio and Narayan, 2015). I use the IHS instead of the standard log transformation to avoid dropping observations with null values.

earlier is not statistically different than the 2003 effect, but post-2003 the effects inversely track the estimated charter entry effects. This figure foreshadows the result in Section 2.5 that charter competition depresses the appraised property values used to compute district local revenues.²⁹ The absence of a pre-trend during the pre-2003 years suggests that this effect is attributable to the increase in charter competition accompanying a lagged AW rating instead of other unobserved trends not handled by my estimation strategy.

Figures 2.4c and 2.4d depict the same test on instructional expenditures and capital outlays. Overall, there is little evidence of pre-trends prior to 2003 providing support for the efficacy of the identification strategy. Further, I find evidence of negative post-2003 effects on instructional spending and noisy, positive effects on capital outlays. Finally, the results for the rest of the outcomes explored in this paper are presented in Figures G.1 through G.7 of Online Appendix G. Overall, these tests support the instrumental variables framework from (2.2) as a plausibly valid estimation strategy.

2.4 Student and Teacher Mobility Responses

Before looking at how charter competition affects the budget and resource allocation among traditional public school districts, it is helpful to assess how charter competition influences student and teacher mobility. There is a mechanical relationship between charter transfers and the sending district’s total enrollment. Losing one student to a charter school will mechanically decrease the sending district’s enrollment by one. If not, then charter transfers likely correlate with other types of either student entry into or exit from the TPSD suggesting that my estimation strategy is unable to fully isolate the effect of charter transfers. To test this directly, I estimate equation (2.2) using “potential enrollment” (i.e., actual enrollment + charter transfers) as the outcome; however, I instead measure charter transfers and “potential enrollment” in levels. Theoretically, this regression should yield a null coefficient if charter competition does not induce any other type of entry into or exit from the district. Indeed, I estimate that the effect of transferring a single student to a charter

²⁹While the effect of lagged AW ratings on property values are the most extreme during the Great Recession, I show in Online Appendix E that the Great Recession does not drive my results.

school on the district’s overall “potential enrollment” is statistically indistinguishable from zero (a 0.099 point estimate with a standard error of 0.235). This is consistent with the idea that charter competition is not inducing additional exit to private schools for example (Chakrabarti and Roy, 2016) or exit from the state.

Table 2.3 displays the effect of a one percentage point increase in the fraction of TPSD students who transfer to charters on a range of mobility outcomes. The table shows the results from baseline OLS estimates of equation (2.1) and my preferred instrumental variables (IV) estimates of equation (2.2). The accompanying first-stage estimates and the weak instrument test are located in the table notes. To better understand what types of students are leaving the districts, columns 1 and 2 respectively present the effect of charter transfers on the inverse hyperbolic sine (IHS) of free/reduced-price lunch (FRL) eligible and special education student enrollment. Because these and other outcomes can take on null values, I use an IHS transformation to give the coefficients a similar interpretation as a log transformation without losing null observations. I estimate that a one percentage point increase in the fraction of charter transfers reduces FRL eligible and special education student enrollment by 6.9 percent and 3.2 percent, respectively.

In response to the overall decline in the number of enrolled students, if districts did nothing to adjust teacher labor supply, then student-teacher ratios would decrease. However, columns 3 and 4 show that districts respond to the decrease in enrollment by reducing the size of the teaching force in lock-step. Student-teacher ratios are preserved by a 3.3 percent reduction in overall teaching staff. Column 5 reveals that this reduction is partly driven by hiring 7.6 percent fewer new teachers.

Finally, columns 6 and 7 show the effect of charter competition on teacher exit and entry between charter schools and TPSDs. As is pointed out by Jackson (2012), because teachers can only move between charters and TPSDs when charters are present, there will be a mechanically positive relationship between charter competition and these measures of teacher mobility. I estimate that a percentage point increase in charter competition increases teacher exit to charters by 9.5 percent and increases teacher entry into TPSDs from charters by 4.7 percent. However, note that these estimates are off of an extremely small base. Specifically, districts at the 99th percentile of the TPSD-charter teacher mobility distribution only lose 1 teacher to charters and also only gain 1 teacher from charters. Overall, the evidence from this table supports a conceptual framework

where districts are competing with charters over students instead of over teachers.³⁰ Further note that across the table, the baseline estimates are qualitatively similar to my preferred specification though are more precisely estimated and are smaller in magnitude.

All three of the instruments generate statistically significant increases in charter entry (see the table note).³¹ For example, a district with a two-period lagged “Academic Emergency” rating after the 1999 policy experiences roughly a 1.5 percentage point increase in the fraction of students attending a charter school compared to an equally-rated district prior to 1999. The table note also provides a weak instrument test. The large F statistic for the excluded instruments show that these regressions do not suffer from weak instruments. Hansen J statistics and p-values for tests of overidentification are provided in Online Appendix Table H.1 for all outcomes explored in this paper. The null hypotheses for these tests are that the instruments are valid. Across the different outcomes, I find suggestive evidence supporting the validity of my instruments.

2.5 District Revenues

With the responses of student and teacher mobility to charter transfers in mind, I now analyze how charter competition influences TPSD revenues. In Ohio, when a student transfers to a charter school, the state funding is still paid to the student’s district of residence as if the student were still enrolled. However, the district is then required to transfer a formula-derived amount of state funding directly to the charter. This transfer is recorded as an expenditure.

Panel A of Table 2.4 presents baseline OLS estimates from equation (2.1) and my preferred IV estimates from equation (2.2) for the IHS of real total, federal, and local revenues.³² Due to the

³⁰In nearly every dimension, employment at a TPSD dominates employment at a charter school. Pay is often better, there is superior job security due to union affiliation, and work hours are often shorter. As a result, it is plausible that TPSDs will not be competing with charters to retain their teachers.

³¹The urban district instrument is only marginally significant in these specifications. For the remaining specifications in Tables 2.4 to 2.6, this instrument is significant at the 1 percent level because I utilize a slightly larger regression sample.

³²Since the treatment is a loss of students to charter schools, in order to avoid a mechanical denominator bias in my results Tables 2.4 through 2.6 present estimates for total revenue/expenditure categories instead of per-pupil

slightly increased sample size as compared to the student and teacher mobility regressions from Table 2.3, the first-stage estimates are even more precisely estimated (see table notes). I find that increasing the fraction of students transferring to a charter school by one percentage point decreases total revenues by 1.8 percent. To put the size of this effect into context, a percentage point increase in the fraction of students transferring to a charter represents an increase of roughly half of a standard deviation (see Online Appendix Table H.2) or about half of the growth in average charter competition from 1998 to 2010.³³

Decomposing total revenues by government level (columns 2 and 3) reveals that charter competition induces revenue losses at both the federal and local levels. However, I estimate that charter competition has little effect on state revenues.³⁴ The lack of impact that charter transfers have on state revenues is not surprising. Because state funding is paid to the sending district and then transferred to the charter as an expenditure, there is no mechanical link between charter transfers and state revenues. However, while technically counted as an expenditure, these charter payments should be thought of as effectively decreasing the state funding available to TPSDs.

The negative effect of charter competition on federal funding is also not surprising. Major federally funded programs contained in the Child Nutrition Act and the Individuals with Disabilities Education Act for example provide per-pupil funding for certain eligible student groups. Distribution of these federal entitlement grants to school districts are formula-based. Thus, when a student funded under any of these federal programs transfers to a charter school the accompanying federal revenues are deducted from the TPSD and are instead allocated to the charter. Recall from Table 2.3 that charter competition decreases the TPSD enrollment of FRL-eligible and special education students. As a result, charter competition should mechanically decrease federal revenues.

To see these channels empirically, I further decompose federal revenue effects in Panel B. Columns 5 and 6 present estimates for two large federal programs and column 7 presents the effect on all other federal revenues. Federal funding is decreasing across both programs.

values.

³³The fraction of students transferring to charters in 2010 was roughly 0.03 among districts in the regression sample.

³⁴For my IV specification, I estimate that a one percentage point increase in charter competition decreases state revenues by a statistically insignificant 1.0 percent.

Unlike the effects of charter competition on federal and state revenues, the negative effect on local revenues is unexpected. I explore potential mechanisms in Panel C by decomposing local revenues into the contribution of local property taxes, school lunch funding, and all other local revenues in columns 8 through 10. While local revenues are decreasing across all three measures, I focus my discussion on local property taxes because they comprise 96.5 percent of local revenues (LSC, 2011). Charter competition can affect local property taxes through two main channels. First, competition can directly decrease appraised property values and, in turn, the base valuation being taxed. Second, charters can decrease the levied millage rates (i.e., one-tenth of one percent) that determine the fraction of the base property values being taxed.

I test these potential mechanisms in columns 11 through 13 of Panel D. In column 11, I present the effect of charter competition directly on the total appraised property values within the TPSD. My estimates suggest that a percentage point increase in the fraction of TPSD students transferring to charters decreases real appraised property values by 2.5 percent. This property value measure aggregates residential, agricultural, commercial, industrial, and mineral properties, as well as other public properties (LSC, 2011).³⁵

Column 12 presents the effect of competition solely on appraised residential property values. Charter competition generates nearly identical percent losses for both total property values and residential property values.³⁶ In column 13, I find that the millage rates tend to decrease as charter competition increases. A simple back-of-the-envelope decomposition reveals that a majority of the decrease in total revenues is driven by the change in total property values.³⁷

Property values are appraised every six years by a county auditor through visual inspection. Every three years, the appraised value is updated using market transaction data and forecasting algorithms to estimate the value of the property (Sullivan and Sobul, 2010). As a result, the decrease in

³⁵Residential and agricultural properties make up 79.1 percent of total property values in 2008 (LSC, 2011).

³⁶In Section 2.3.2 and Online Appendix E, I present evidence that these depressed housing values are not driven by the Great Recession.

³⁷Local revenues (LR) are calculated using $LR = \frac{\text{millage}}{1000} \times \text{Property Values}$. The marginal effect of charter competition on property values relative to the average is roughly a \$10,200,000 decrease. Relative to average millage rates, a decline in property values of this magnitude decreases local revenues by roughly \$300,000. Conversely, relative to average property values, a 0.219 drop in the millage rate decreases local revenue by approximately \$90,000.

appraised residential property values could reflect true depreciation of underlying housing values as well as changes in the appraisal process. To rigorously assess the housing capitalization of charter schools I would need parcel-level housing sales data as in Imberman et al. (2016).

Still, the negative effect of charter competition on appraised housing valuation is unexpected. If there is an excess demand for schooling, one might suppose that additional schools would positively capitalize into local housing values. However, competing forces plausibly exist. There is evidence in the literature, and I find evidence in Ohio that charter school transfers generate fiscal stress for the sending TPSDs (Bifulco and Reback, 2014; Arsen and Ni, 2012b). In the spirit of Bifulco and Reback (2014), I estimate that for each student transferring to a charter school, the sending TPSD will on average save \$4,027 in variable costs from not having to educate the transferring student.³⁸ However, this savings only accounts for roughly two-thirds of the state revenue reductions accompanying each charter transfer yielding a net loss. If fiscal stress lowers *perceived* TPSD quality, then housing prices within the TPSD would also likely decrease to reflect these perceptions (Black, 1999). Thus, the direction of the effect of charter competition on housing prices depends on which force dominates.

There are several reasons to believe that the negative pressures could plausibly dominate in Ohio. First, if charter quality is low, then opening new charters may not generate positive housing capitalization. The Ohio Department of Education places relatively few restrictions on the eligibility of charter school authorizers and implements limited restrictions to ensure quality control.³⁹ As a result, the quality of the average Ohio charter school may be lower than in other states and may be negatively capitalized into housing values.

Second, positive effects likely apply generally across large geographic areas, while negative pressures are likely district-specific. Neither brick-and-mortar nor digital charter schools have geographic

³⁸I separate variable costs into rough measures related to student- and teacher-related per-unit expenditures (e.g., pupil services vs. teacher salaries). I then convert my measure of per-teacher variable costs to per-pupil variable costs by multiplying by 0.441, my IV estimate for the effect of a single charter transfer on total employed teachers. My final variable costs measure then combines the per-pupil variable costs with this measure of per-teacher variable costs converted to per-pupil units.

³⁹Ohio has been referred to as the “Wild, Wild West” of charter schools by organizations such as the National Association of Charter School Authorizers (O’Donnell, 2014).

enrollment boundaries allowing the potential capitalization gains to spread across district borders. As a result, charter presence may provide a general housing value premium across districts and only a negligible relative premium between nearby districts. Conversely, any housing valuation penalties due to charter-related fiscal stress on local TPSDs would show up as differences in relative housing values between districts. This is relevant for my setting because if new charter entry in a given district improves school choice for parents across all districts in the commuting-zone, this general effect will be subsumed by my commute-zone-by-year fixed effects.⁴⁰

Third, housing prices within the subset of districts identifying the LATE might be more sensitive to charter loss than the average district. In an unpublished working paper, [Buerger \(2014\)](#) finds negative effects for charter competition on housing prices in poorer neighborhoods. In Section 2.3.2, I described how my IV strategy provides LATE estimates for low-performing school districts and in Ohio the most economically disadvantaged districts also tend to be the lowest performing. Thus, consistent with Buerger’s findings, my negative housing capitalization effects are local to poorer neighborhoods.

There is additional evidence in the literature that charter competition may put downward pressure on residential property values. [Imberman et al. \(2016\)](#) study housing sales price responses to charter competition in the Los Angeles Unified School District (LAUSD) and find no effect on average. However, upon restricting attention to houses outside of the LAUSD, [Imberman et al. \(2016\)](#) find that additional charters entering within a household’s TPSD boundary have a negative effect on housing sales prices.

As students transfer to charter schools, the only mechanical effect on a district’s budget is the required transfer of state funding to the charter. A tempting yet spurious conclusion would be to infer that charters only reduce the overall size of the budget through this mechanical increase in charter-transfer expenditures. However, the main takeaway from Table 2.4 is that charter competition places additional fiscal stress on TPSDs as federal revenues decline from “at-risk” student transfers and local revenues are lost from depressed residential housing valuations.

⁴⁰Regressing the main IV specification in (2.2), but substituting year fixed effects for commute-zone-by-year fixed effects generates statistically indistinguishable estimates. As a result, this concern is not likely a first-order issue.

2.6 Collectively Bargained Teacher Contracts

In Ohio, instructional expenditures alone make up over half of total expenditures. Further, Ohio is a heavily unionized state and teacher pay is determined through collective bargaining between TPSDs and teachers' unions. As a result, before assessing the effect of charter competition on general district resource allocation, I first highlight the effect of charter competition directly on collectively bargained contract outcomes.

Table 2.5 presents the results from estimating the OLS and IV specifications respectively from equations (2.5) and (2.6) for several collectively bargained contract outcomes. Column 1 reports that a percentage point increase in the fraction of TPSD students attending a charter decreases real entry-level salaries by 0.2 percent though this effect is not statistically significant. Column 2 presents the effect of charter transfers on top-level salaries. In both specifications, I estimate that competition does not affect top-level salaries. Recall that in the SERB data, top-level salaries are often coded as the lowest value for which an additional year of experience has no effect on salary.⁴¹ As a result, in column 3, I present results for top-level salaries imputed from ODE teacher-level data using the algorithm detailed in Online Appendix B.⁴² For my preferred estimates, I find that a percentage point increase in the fraction of students transferring to charter schools decreases imputed top-level salaries by 1.0 percent. Relative to the average imputed top-level salary, this translates to a \$599 annual salary decrease for teachers at the top of their pay scale.

While these decreases may appear an individual teacher's perspective they reflect non-negligible instructional expenditure reductions for the district as a whole. The average TPSD employs 176 teachers implying that a percentage point increase in charter transfers induces instructional cost reductions up to roughly \$100,000.⁴³ To further put these effect sizes into context with the literature, Hoxby (1996) estimates that initial unionization increases subsequent teacher salaries by 5 percent. Using this estimate as a baseline for the union wage premium, I estimate that a percentage point

⁴¹For example, in Online Appendix Table A.2 the top salaries for non-degree, BA, and MA teachers would be coded respectively as \$29,265, \$40,401, and \$46,616.

⁴²Specifically, this measure is created by taking the maximum salary from the 15-20th imputed pay scale steps using the method in Online Appendix B and filling in missing values with SERB top-level contract outcomes.

⁴³Potential cost reduction is calculated assuming that all teachers are at the top of the pay scale.

increase in the fraction of charter transfers erodes about 20 percent of the union wage premium for the most experienced teachers.

Column 4 characterizes whether charter competition changes the slope of the negotiated salary profile. In these regressions the outcome is the difference between real top- and entry-level salaries divided by the number of steps it takes to reach a top-level salary. I estimate that charter competition flattens the pay scale, though the effect is only significant for the baseline OLS specification. In addition to affecting the negotiated salaries directly, charter competition could also change the length of time it takes to ascend the pay scale ladder. Column 5 shows the effect of charter competition on the number of steps required to fully ascend the pay scale. While marginally statistically significant, these estimates represent economically insignificant increases seeing that on average teachers face 15 pay scale steps.

Next, I estimate the effect of charter competition on the entire negotiated salary distribution rather than solely focusing on the top and bottom of the pay scale. This analysis assesses the degree to which the previous negative salary effects generalize across the rest of the pay scale. Even though SERB-provided data only include salary information for entry and top pay scale steps, I approximate the negotiated salaries for each intermediate step using teacher-level wage data to estimate equation (B.1) in Online Appendix B.

Figure 2.5 presents the results from estimating equations (2.5) and (2.6) for each imputed pay scale step. The baseline OLS estimates show almost no response across the pay scale. However, my preferred IV estimates suggest that both entry-level and top-level salaries decrease in the presence of charter competition. I estimate that a percentage point increase in the fraction of charter transfers decreases collectively bargained entry-level salaries by 1.8 percent and top-level salaries by 2.3 percent. Interestingly, charter competition has almost no statistically significant effect across the middle of the pay scale distribution. One explanation is that a majority of the teaching staff is comprised of new teachers at the bottom of the pay scale and veteran teachers who have already reached the top-level salaries. By decreasing salaries for these two groups, districts would be able to enjoy the largest savings.

Overall, finding that charter competition decreases collectively bargained salaries may seem un-

expected, especially considering the positive effects found in non-union settings by [Taylor \(2006, 2010\)](#) and [Jackson \(2012\)](#). However, there are a few ways to frame these findings. First, because charter schools often pay lower salaries to their teachers ([Arsen and Ni, 2012a](#)), competition over students may put downward pressure on TPSD-negotiated salaries to match charter salaries.

Second, this could be a story of union/TPSD negotiating power. Recall [Table 2.3](#) showed that as districts lose students to charter schools, the size of the teaching labor force reduces in lock-step. As a result, charter competition could give TPSDs leverage in contract negotiations. If TPSDs cannot decrease teacher salaries and reallocate resources to prevent student transfers, then the size of the teacher labor force represented by the unions will drop. If the threat of downsizing provides leverage for TPSDs, then the loss of union bargaining power may reduce the artificially high union monopoly wage.

2.7 Resource Allocation

[Table 2.6](#) presents the effect of charter competition on district expenditures. Column 1 provides an estimate of the financial burden placed on TPSDs from charter transfers. Increasing the fraction of students transferring to a charter by one percentage point, increases the amount of money transferred to charters by roughly \$200,000. On average, a percentage point of a district's potential enrollment is about 30 students, which equates to roughly \$6,600 transferred per student, approximately the baseline formula amount in 2010.

To see how total expenditures are influenced by charter competition beyond the mechanical charter-transfer increase, column 2 presents the IHS of total expenditures after netting out any payments to charter schools. IV estimates suggest that total expenditures fall by 1.7 percent which matches up closely with the estimated 1.8 percent decrease in total revenues from [Table 2.4](#).

Columns 3 through 5 present estimates for the effect of charter competition on the allocation of remaining district resources. I show above that increased charter transfers decrease both negotiated teacher salaries as well as the size of the overall teaching force. As a result, we should expect to see fewer resources spent on teaching expenditures in the presence of charter competition. Indeed,

IV estimates suggest that TPSDs facing a percentage point increase in the fraction of students transferring to charters spend 2.3 percent less on instructional expenditures. Curiously though, these districts spend 7.3 percent more on capital outlays while spending 2.8 percent less on all other expenditures. To explore the mechanism driving the surprising increase in total capital outlays, columns 6 and 7 respectively present estimates for the IHS of new construction capital outlays and all other capital outlays. Increases in capital outlays are driven by new construction expenditures (a 11.3 percent increase). These effects are robust across a variety of alternative measures of charter competition (see Online Appendix Table C.1).

2.8 Discussion

There are several reasons why TPSDs might allocate resources toward capital outlays in response to charter competition. First, suppose principals believe that capital outlays enhance subsequent student performance and school ratings.⁴⁴ Because parents factor school ratings information into student enrollment decisions (Cullen et al., 2006; Hastings and Weinstein, 2008; Hanushek et al., 2007), charter competition provides incentives for district administrators to allocate resources to areas that generate gains to student achievement and boost subsequent enrollment. Thus, if capital outlays improve student performance, then charter competition creates incentives for TPSDs to boost student achievement in ways predicted by traditional school choice theory (see Friedman, 1955, 1997; Hoxby, 2003a,b).

Second, the literature provides no clear suggestion for the ideal combination of school expenditures to optimize student achievement and subsequent school ratings given a fixed budget.⁴⁵ Thus, it

⁴⁴The literature is mixed regarding whether capital spending resulting from narrowly approved local capital bond referenda affect subsequent student achievement. Martorell et al. (2015) find precisely estimated null effects, while Cellini et al. (2010) and Hong and Zimmer (2016) provide evidence that capital outlays can have positive achievement effects several years after the bond passage.

⁴⁵“Unfortunately, identification of truly exogenous determinants of ... resource allocations ... is sufficiently rare that other compromises in the data and modelling are frequently required. These coincidental compromises jeopardise the ability to obtain clean estimates of resource effects and may limit the generalisability of any findings” (Hanushek, 2003, pg. 83).

is plausible that district administrators may also be unsure how to boost achievement through resource allocation. If administrators believe that parents value facility condition, then resources may be channeled to capital outlays where each dollar spent is clearly linked to visible facility improvement regardless of subsequent student achievement. This type of allocation is consistent with qualitative survey evidence on how principals in D.C. try to insulate against charter transfers. “The physical appearance of school buildings was said to have the greatest impact on enrollment trends... We noted that principals did not tend to focus on test scores or academic achievement in their lists of attributes that parents sought when selecting schools” (Sullivan et al., 2008, p. 20). If administrators are spending money on capital because parents value it directly (Cellini et al., 2010) and not because administrators believe it improves student performance, then in this scenario, charter competition does not necessarily improve TPSD achievement efficiency. However, competition is still efficiency-enhancing in the sense that it causes districts to spend in areas valued directly by parents.

Finally, allocation toward capital outlays may be the result of information acquisition costs. Suppose parents value student achievement, but signals of school quality vary in their acquisition costs. While motivated parents will seek out already available school quality information when provided school choice (Lovenheim and Walsh, 2014), low-cost albeit noisy signals of school quality such as facility condition potentially inform even the time-constrained or less motivated parents. In this case, TPSDs may respond to competition by allocating resources toward these salient signals regardless of the impact on subsequent achievement. This behavior is again consistent with qualitative survey evidence from D.C. principals.⁴⁶ If parents only value facility condition because they mistakenly infer information about a school’s potential education production and if capital outlays do not affect achievement, then charter competition actually exacerbates the misallocation of resources. Under this framework, simple policies can potentially correct this incentive structure to instead encourage TPSDs to allocate resources in ways that improve student achievement.⁴⁷

⁴⁶ “According to our sample, it appears that most of the changes that schools are making in order to attract more students [from charters] have more to do with services for parents and the image of the school than with improving the educational attainment of students” (Sullivan et al., 2008, p. 21).

⁴⁷ Reducing the cost to obtain school achievement information may help correct competitive incentives (Hastings and Weinstein, 2008). For example, adding simple statistics comparing academic ratings between the schools in the parent’s choice set onto any required school choice form could provide such salient and relevant school quality

These three scenarios highlight that charter competition has the potential, but is not guaranteed to encourage TPSDs to reallocate resources in ways that enhance student achievement as predicted by economic theory. Disentangling the mechanisms underlying why charter competition causes TPSDs to reallocate resources is a rich area for future work. Overall, because facility condition is likely valued by parents, as measured by positive housing capitalization (Cellini et al., 2010), I interpret my results as evidence that charter competition causes TPSDs to allocate resources toward areas that are likely valued directly by parents, but that do not necessarily improve student achievement.

2.9 Conclusion

The charter school movement is one of the fastest growing education reforms in the United States. Charters are designed in part to inject competition into the education market to boost TPSD student achievement. There is a large and mixed literature assessing the effect of charter competition directly on TPSD student achievement (Epple et al., 2015). However, little attention has been given to potential mechanisms. I fill this gap in the literature by analyzing the effect of charter competition on TPSD revenue and resource allocation. I also pay special attention to how charter competition affects collectively bargained teacher compensation in a strongly unionized state. To accomplish this, I exploit both the long charter approval process as well as plausibly exogenous variation in policies that determine the location and timing of Ohio charter entry in an instrumental variables framework. I collect and merge together several datasets, including the universe of Ohio public school teachers' union contracts as well as district-level charter school transfer information from Ohio Department of Education financial reports.

I find that charter competition directly decreases TPSD revenues in excess of the mechanical loss of state resources due to lower enrollment. As vulnerable student populations transfer to charters, TPSDs lose the federal funding designated to help educate these students. Further, I show that charter competition indirectly decreases TPSD revenues by depressing the appraised value of residential properties, thus lowering the base from which local revenues are taxed. To help mitigate the erosion of TPSD local revenues, states could consider providing countervailing aid to districts

information.

facing heavy charter competition.⁴⁸

Other key findings include that charter competition causes districts to negotiate lower unionized teacher salaries, spend less on instructional and other expenditures, and spend more on new construction expenditures. Determining whether this type of charter-driven resource allocation improves student achievement is an important area for future work.

⁴⁸For example, districts facing heavy charter transfers in New York receive state transitional aid designed to mitigate negative fiscal impacts (Bifulco and Reback, 2014).

Table 2.1: Descriptive Statistics by Charter Competition Categories

	Full Sample	Charter Transfers		
		None	Some	Top Quartile
	(1)	(2)	(3)	(4)
District Characteristics				
Potential Student Enrollment per District	2,884 (4,951)	2,364 (3,849)	3,026 (5,202)	4,724 (9,086)
Teachers per School	176 (313)	147 (256)	184 (326)	283 (570)
Real Total Property Value (in thousands)	427,855 (729,160)	329,178 (506,929)	454,770 (776,704)	616,449 (1,257,359)
Fraction Black Students	0.054 (0.133)	0.041 (0.116)	0.057 (0.137)	0.134 (0.217)
Fraction FRL Students	0.202 (0.155)	0.148 (0.118)	0.216 (0.161)	0.323 (0.182)
Fraction Student Transfers to Charter	0.014 (0.023)	—	0.018 (0.024)	0.044 (0.037)
Total Expenditures (in thousands)	33,876 (67,749)	23,651 (42,098)	36,665 (72,955)	61,227 (131,343)
Instructional Spending (in thousands)	16,509 (32,033)	12,046 (22,136)	17,727 (34,139)	28,576 (60,771)
Capital Outlays (in thousands)	3,780 (9,904)	2,568 (4,487)	4,110 (10,902)	7,075 (18,574)
N	8,474	1,816	6,658	1,664
Contract Characteristics				
Entry-Level Salary	34,816 (5,097)	33,336 (4,641)	36,875 (4,986)	36,212 (4,603)
Top-Level Salary	57,754 (11,348)	54,696 (10,358)	62,013 (11,320)	61,657 (10,521)
N	15,948	9,286	6,474	1,296

Notes: Means and standard deviations (in parentheses) are presented. Contract-by-Salary-Track observations represent a given district-by-contract-start-year-by-education-level cell for either entry or top levels of experience. Column 1, provides information on all district-years (contracts) missing none of the variables in the table. Columns 2 and 3 further conditions on whether a given district has transferred no students and any students to charter schools, respectively for the given year/contract-start-year. Column 4 only includes districts/contracts where the district faces the top quartile of charter competition in the given year.

Table 2.2: Annual District Charter Eligibility and New Charter Entry

	<i>Urban 8/21</i> Districts	Ratings (Emergency or Watch)	Total	# New Charters Opening Next Year
	(1)	(2)	(3)	(4)
1996	0	0	0	0
1997	18	0	18	13
1998	18	0	18	25
1999	31	0	31	17
2000	31	21	52	33
2001	31	3	34	47
2002	31	6	37	34
2003	18	60	78	90
2004	18	31	49	69
2005	18	21	39	21
2006	18	4	22	26
2007	18	8	26	20
2008	18	6	24	14
2009	18	7	25	37
2010	18	8	26	32
2011	18	4	22	—

Notes: The table shows the number of districts eligible for charters to begin the process of opening in each given year. Column (1) shows the number of districts eligible based on urbanicity, i.e., whether the district is in Lucas county, or is one of the *Big 8* or *Urban 21* districts during a year that policy allows charter entry. Column (2) presents the number of districts eligible for new charter entry in the given year based exclusively on eligibility determined by academic ratings during the previous year. If a district is eligible for charter entry based on the criteria in both Columns (1) and (2), the district is only counted in Column (1). Column (3) gives the total number of eligible districts. Column (4) presents the number of new charter schools that will open in the subsequent year due to eligibility in the given year. For example, 25 charter schools opened in 1999 based on district eligibility during 1998. In 2011 there were 608 non-charter school districts and 355 charter schools.

Table 2.3: Effect of Charter Transfers on Student and Teacher Mobility

	IHS of Student Count			IHS of Teacher Count			
	FRL Eligible (1)	Special Education (2)	Stu/Tch Ratio (3)	Total (4)	First Year (5)	Exit to CS (6)	Enter from CS (7)
Fraction of Charter Transfers $\times 100 - \text{OLS}$	-0.048*** (0.009)	-0.016*** (0.003)	0.003 (0.022)	-0.020*** (0.003)	-0.031*** (0.006)	0.043*** (0.009)	0.025*** (0.007)
Fraction of Charter Transfers $\times 100 - \text{IV}$	-0.069*** (0.015)	-0.032*** (0.007)	-0.008 (0.061)	-0.033*** (0.006)	-0.076** (0.031)	0.095*** (0.018)	0.047* (0.026)

Notes: N= 8,405 district-year observations. Standard Errors in parentheses are clustered by district. See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). First-stage estimates (and standard errors) for excluded instruments are: Post $1999_{t-2} * \mathbb{1}(\text{Acad. E.})_{t-2} = 1.440^{***}$ (0.518); Post $2002_{t-2} * \mathbb{1}(\text{Acad. W.})_{t-2} = 2.569^{***}$ (0.562); and $t - 1$ Char. Elig. (Urban 8/21)=0.334 (0.720). The F statistic for excluded instruments is 8.021***. This table reports OLS and 2SLS estimates of the effect of charter competition on different forms of student and teacher mobility. The endogenous variable is the fraction of the district's potential enrollment that instead transfers to a charter school times 100. Regressions in Panel A include district and commute-zone-by-year fixed effects. Regressions in Panel B further instrument for charter competition using the instruments and additionally include the main effects for the instruments and $t - 1$ Academic Watch/Emergency indicator variables. Each column provide the results of a separate regression. See Online Table H.2 for the mean of each dependent variable and Online Table H.1 for tests of overidentification.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 2.4: Effect of Charter Transfers on District Revenues

Panel A: IHS of Total Revenues	Total	Federal	Local
	(1)	(2)	(3)
Fraction Charter Transfers $\times 100$ – OLS	-0.007*** (0.002)	-0.027*** (0.005)	-0.021*** (0.003)
Fraction Charter Transfers $\times 100$ – IV	-0.018*** (0.005)	-0.041*** (0.007)	-0.034*** (0.006)
Panel B: IHS of Federal Revenues	Child Nutrition Act	Disabilities Act	Other
	(5)	(6)	(7)
Fraction Charter Transfers $\times 100$ – OLS	-0.033*** (0.011)	-0.065*** (0.022)	-0.010** (0.005)
Fraction Charter Transfers $\times 100$ – IV	-0.045*** (0.016)	-0.068 (0.073)	-0.003 (0.010)
Panel C: IHS of Local Revenues	Property Tax	School Lunch	Other
	(8)	(9)	(10)
Fraction Charter Transfers $\times 100$ – OLS	-0.020*** (0.003)	-0.060*** (0.009)	-0.008 (0.007)
Fraction Charter Transfers $\times 100$ – IV	-0.028*** (0.005)	-0.062*** (0.016)	-0.036** (0.015)
Panel D: Property Tax Decomposition	IHS of Property Value		
	Total	Residential	Millage
	(11)	(12)	(13)
Fraction Charter Transfers $\times 100$ – OLS	-0.023*** (0.003)	-0.017*** (0.003)	-0.002 (0.086)
Fraction Charter Transfers $\times 100$ – IV	-0.025*** (0.004)	-0.026*** (0.005)	-0.219* (0.121)

Notes: N= 11,449 district-year observations. Standard Errors in parentheses are clustered by district. First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{t-2} * 1(\text{Acad. E.})_{t-2} = 2.306^{***}$ (0.691); Post 2002 $_{t-2} * 1(\text{Acad. W.})_{t-2} = 3.573^{***}$ (0.649); and $t - 1$ Char. Elig. (Urban 8/21) = 2.471*** (0.927). See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). The F statistic for excluded instruments is 10.315***. This table reports OLS (see equation (2.1)) and 2SLS (see equation (2.2)) estimates for the effect of charter competition on district revenues. The endogenous variable is the fraction of the district's membership attending charter schools times 100. Each cell provides the result of a separate regression. See Online Table H.2 for the mean of each dependent variable and Online Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table 2.5: Effect of Charter Transfers on Collectively Bargained Contracts

	Log of Real Salary			Other Contract Outcomes	
	Entry	Top	Imputed Top	Slope	# Payscale Years
	(1)	(2)	(3)	(4)	(5)
<i>t</i> -1 Fraction Charter Transfers $\times 100$ – OLS	-0.004*** (0.001)	-0.001 (0.001)	-0.001 (0.001)	-8.647*** (3.321)	0.148** (0.072)
<i>t</i> -1 Fraction Charter Transfers $\times 100$ – IV	-0.002 (0.002)	-0.001 (0.002)	-0.010** (0.004)	-6.910 (5.939)	0.168 (0.104)

Notes: $N = 13,930$ contract observations. Standard Errors are in parentheses and are clustered by district. First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{\tau-3} * \mathbb{1}(\text{Acad. E.})_{\tau-2} = 3.877^{***}$ (1.427); Post 2002 $_{\tau-3} * \mathbb{1}(\text{Acad. W.})_{\tau-2} = 3.403^{***}$ (0.750); and $\tau - 2$ Char. Elig. (Urban 8/21) = 2.481*** (0.859). See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). The F statistic for excluded instruments is 10.285***. This table reports OLS (2.1) and 2SLS (see equation (2.2)) estimates of the effect of charter competition on negotiated pecuniary and non-pecuniary contract outcomes. *Imputed Top* in Column (3) is calculated by using the maximum salary from the 15th through 20th imputed payscale estimates as calculated from the procedure detailed in Appendix B and filling in missing values with SERB top-level contract outcomes. The endogenous variable is the fraction of the district's total enrollment lost to any charter schools, i.e., # transferred / (# in district + # transferred). See Online Table H.2 for the mean of each dependent variable and Online Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table 2.6: Effect of Charter Transfers on District Expenditures

	IHS of Expenditure			IHS of Capital Outlays		
	Charter Payments (100,000s)	Total (Net of Charter Payment)	Instruction	Capital Outlays	Other	New Construction
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Charter Transfers $\times 100 - \text{OLS}$	1.144*** (0.257)	-0.007*** (0.002)	-0.021*** (0.004)	0.071*** (0.016)	-0.021*** (0.003)	0.143*** (0.048)
Fraction Charter Transfers $\times 100 - \text{IV}$	1.997*** (0.432)	-0.017*** (0.006)	-0.023*** (0.004)	0.073** (0.034)	-0.028*** (0.004)	0.113 (0.114)
						0.001 (0.019)

Notes: N= 11,449 district-year observations. Standard Errors in parentheses are clustered by district. See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{t-2} * \mathbb{1}(\text{Acad. E.})_{t-2} = 2.306^{***}$ (0.691); Post 2002 $_{t-2} * \mathbb{1}(\text{Acad. W.})_{t-2} = 3.573^{***}$ (0.649); and $t - 1$ Char. Elig. (Urban 8/21)=2.471*** (0.927). The F statistic for excluded instruments is 10.315***. This table reports OLS and 2SLS estimates of the effect of charter competition on different forms of teacher mobility. The endogenous variable is the fraction of the district's membership attending charter schools times 100. Each regression includes district and commute-zone-by-contract-start-year fixed effects. Each Panel and column provide the results of a separate regression. See Online Table H.2 for the mean of each dependent variable and Online Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

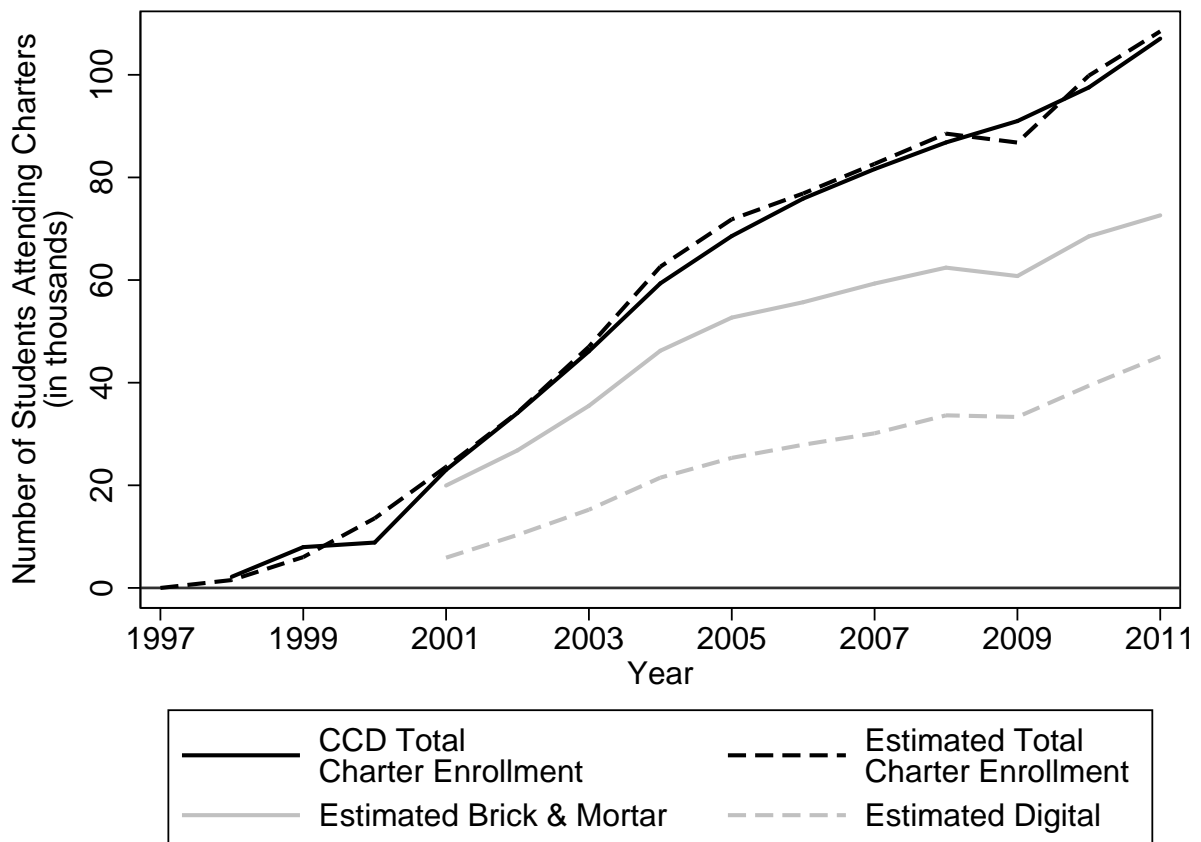


Figure 2.1: Charter Growth

Notes: Data for students attending any charter from 1998-2001 are from CCD payments to charter schools, divided by the baseline amount transferred per student in that year. From 2002 on, charter schools transfers are the sum of student transfers to digital and brick and mortar charters collected from District Foundation Settlement Reports from the ODE. Actual charter counts are the total number of students attending charter designated LEAs from the CCD LEA Universe Survey.

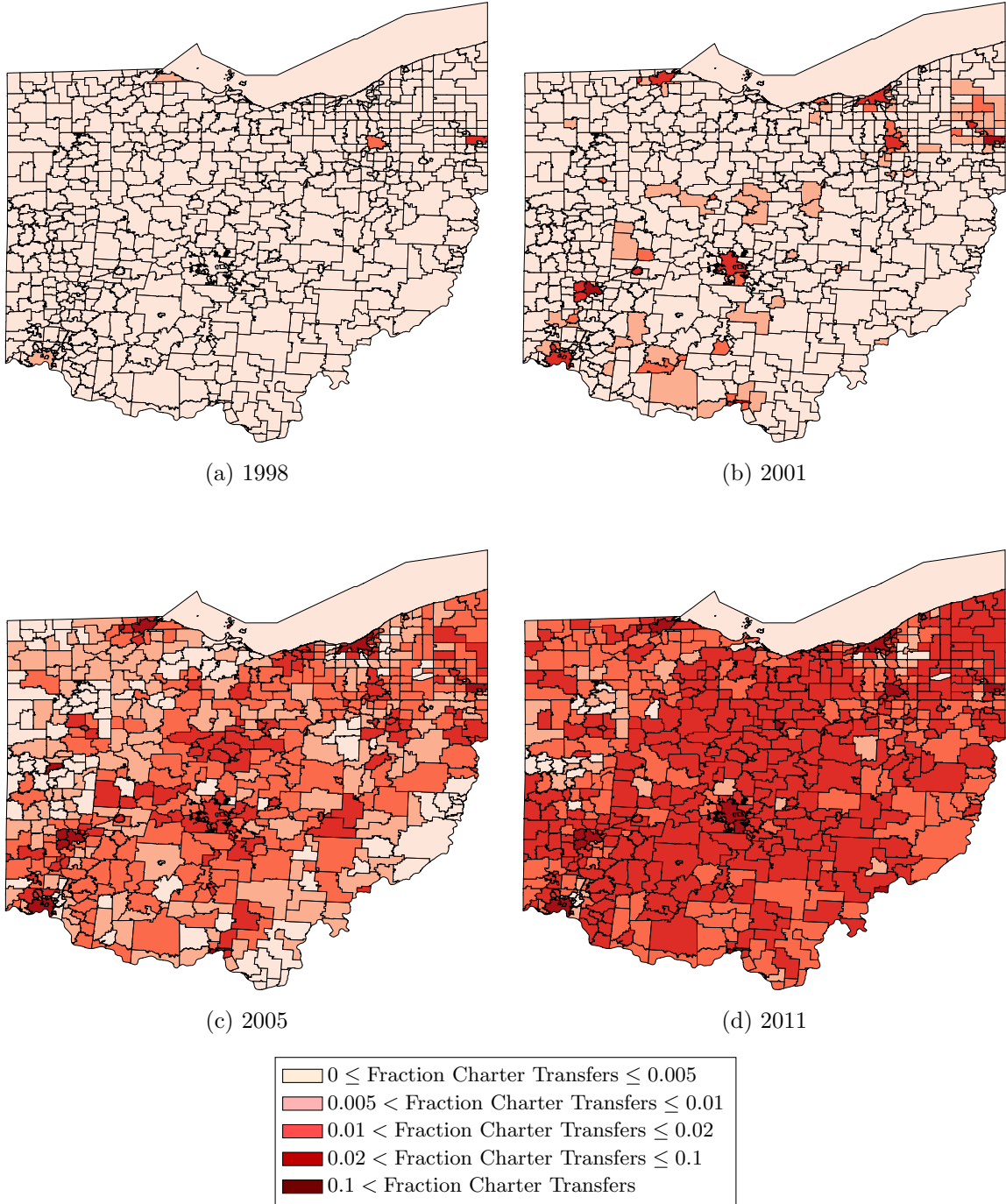


Figure 2.2: Ohio Charter Entry for 1998, 2001, 2005, and 2011

Notes: Ohio district boundaries are plotted and each district is shaded based on the fraction of potential student enrollment that instead transfers to charters for the given year, $\frac{\# \text{ Transferring to Charter}}{\# \text{ Enrolled in District} + \# \text{ Transferring to Charter}}$. See Section 2.2.2 for a detailed explanation for how the # of transferring students is calculated.

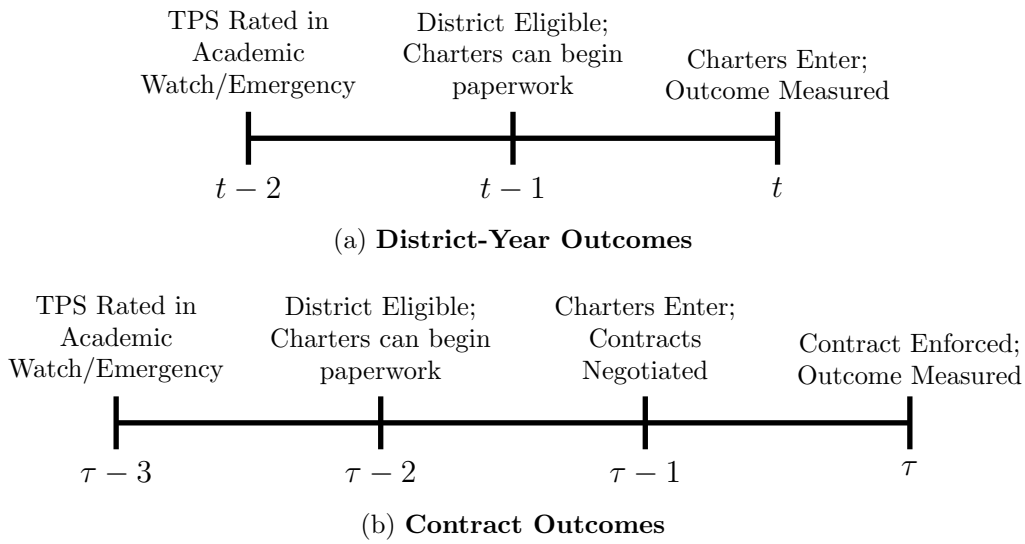


Figure 2.3: Charter Entry Timeline for District-Year and Contract Outcomes

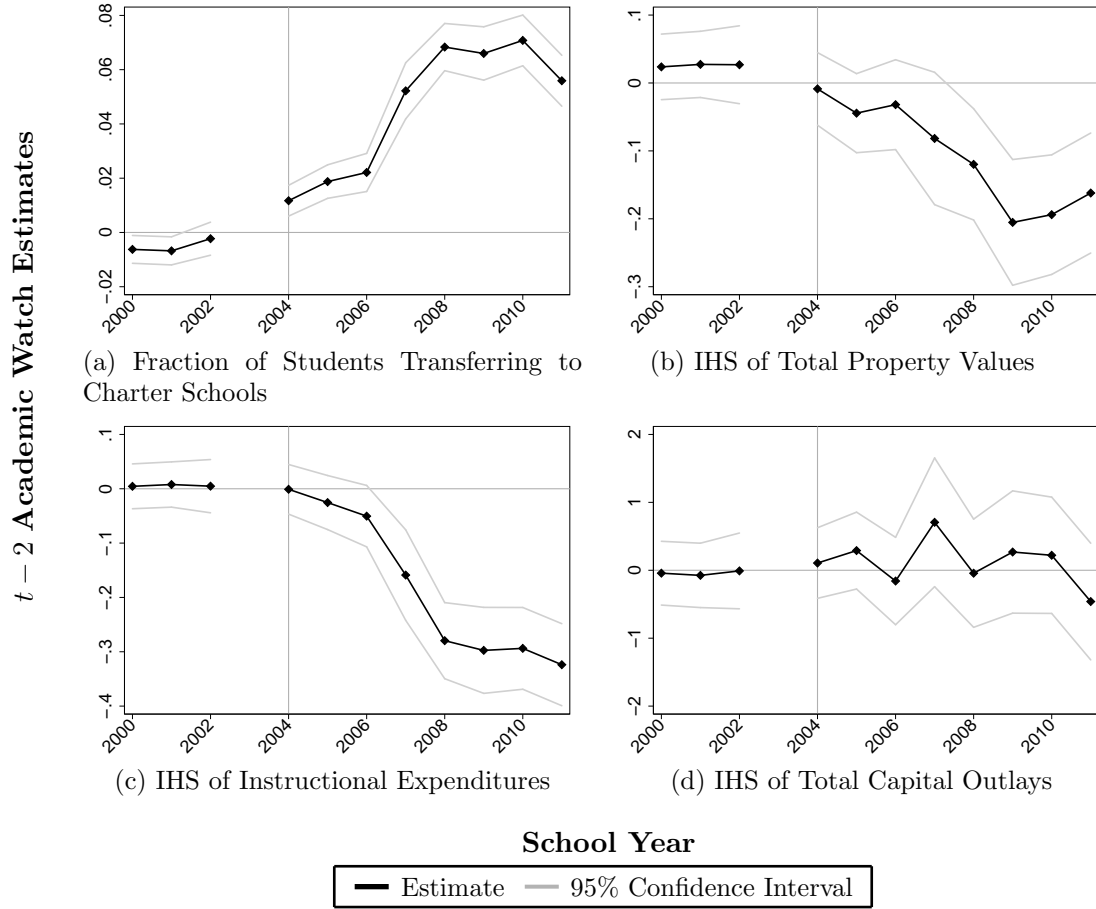


Figure 2.4: Lagged “Academic Watch” Event Study: Various Outcomes

Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

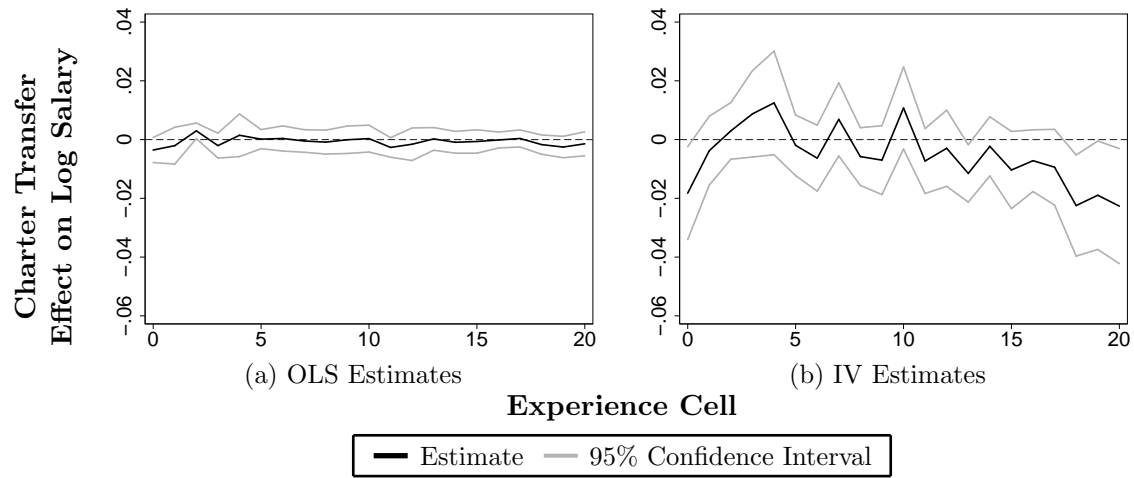


Figure 2.5: Full Imputed Salary Schedule Estimation

Notes: The figure presents the OLS/IV estimates of a 0.01 increase in the fraction of potential district enrollment that transfers to charters on imputed salaries for each experience pay scale step using regressions from equations (2.5)/(2.6). See Online Appendix B for an explanation of the pay scale step imputation method. Each cell presents the results from a separate regression.

Chapter 3

Task-Specific Experience and Task-Specific Talent: Decomposing the Productivity of High School Teachers

3.1 Introduction

How should principals allocate teachers to courses so as to maximize the teachers' contribution to student achievement? The optimal course assignment depends on teachers' existing comparative advantages in different courses or classroom environments, as well as the extent to which current assignments will increase teachers' future productivity (or the principal's information about such productivity). However, the large literature on teacher value-added and the returns to teaching experience (discussed below) has focused primarily on estimating variation in teacher productivity that is assumed (or restricted) to be common to all course or grade assignments. If this assumption is true, then any allocation of existing teachers with fixed course loads will feature the same distribution of value-added contributions to achievement. However, if this assumption is false, then improving the mechanism by which teachers are assigned to courses may yield significant gains at potentially low cost ([Jacob and Rockoff \(2011\)](#)).

To see this, suppose first that teachers have pre-determined comparative advantages for particular subjects or difficulty levels. Then course swaps among teachers could produce efficiency gains if both teachers move toward their relatively more effective courses. Furthermore, if principals cannot ascertain teachers' relative strengths at the time of hire, then the optimal assignment strategy might involve rotating teachers across several different classroom contexts early in their careers so as to produce information about the courses the teachers will be particularly effective at teaching. Permanent subject-specific skill might exist, for example, if a teacher's choice of undergraduate major leads to a deeper understanding of the content in a particular subject (e.g. Physics rather than Biology). Permanent level-specific skill might exist, for example, if a teacher's natural charisma or sense of humor leads to strong classroom control skills that are comparatively more important in the remedial or basic level courses where students may tend to be less engaged.

Now suppose instead that task-specific skill is primarily learned through experience rather than predetermined prior to the time of hire. Then rotating the classroom environments to which teachers are assigned will waste a component of each teacher's skill, and slow each teacher's progress toward his/her full potential. Subject-specific experience might be important, for example, if a teacher's knowledge of the subject content deepens with each opportunity to teach it. Track- or level-specific

experience might also be significant if the appropriate pace at which to deliver content depends on student skill and is slowly calibrated with practice. In addition, experience teaching a certain subject-level combination (e.g. honors biology) might be particularly valuable if it allows teachers to hone particular lectures over time that would be inappropriate for either a different level or a different subject.

More generally, knowledge of the importance of task-specific talent and task-specific experience is essential for any employer wishing to maximize the productivity of his/her workforce. For tasks with larger potential experience gains and smaller variance in task-specific innate talent, the key to a productive workforce is employee retention: the optimal strategy is to keep employees of all talent levels at their originally assigned tasks to benefit from experience. Conversely, for tasks featuring smaller experience gains and a larger variance in task-specific talent, the optimal strategy is to lay off or reassign low performing workers in an attempt to either improve general worker skill or identify superior worker-task matches.

Thus, in this paper we introduce a method for decomposing worker productivity into components relating to general talent, task-specific talent, general experience, and task-specific experience. Our decomposition requires data featuring (1) signals (possibly noisy) of individual workers' task-specific output, (2) histories of worker task assignments, and (3) frequent rotation of workers across tasks. We implement our method using the context of high school teachers, in which tasks consist of teaching particular subjects in particular tracks (difficulty levels).¹

Specifically, we use administrative panel data from the North Carolina Education Research Data Center (NCERDC) to decompose teacher effectiveness at improving student achievement into (1) a set of permanent components capturing general talent, subject-specific talent, level-specific talent, and subject-level specific talent, and (2) a set of functions capturing returns from general experience, subject-specific experience, level-specific experience, and subject-level-specific experience. The data track teachers and students in the universe of public high schools in North Carolina from 1997-2009. Critically, the data feature over 24,000 within-teacher changes in subject assignments

¹Throughout the paper below, we use the term “task” to refer to a subject-level combination, while we use the term “context” more generally to refer to particular characteristics or features of the classroom environment, which include but are not limited to the subject and level.

and over 18,000 changes in academic-level assignments. Such rich data permit estimation of an education production function that features general, subject-specific, level-specific, and subject-level-specific experience profiles as well as a full set of school-teacher-subject-level fixed effects. The flexibility of our model allows us to control for many potential biases that might otherwise accompany endogenous course assignment decisions. We then use our results to project the potential student achievement gains that could be reaped by better utilizing knowledge about course-specific experience and skill relative to the course assignment patterns observed in the data.

Myriad papers have estimated education production functions featuring both teacher fixed effects and a common experience profile. The bulk of the evidence suggests that the standard deviation of permanent teacher quality is between .1 and .2 test score standard deviations at both the primary or secondary school levels.² Similarly, the existing literature suggests that while teachers tend to improve with experience by around .05 test score standard deviations in their first year, another .03 to .05 over the next couple of years, and another .03 to .05 over the next several years, with the profile for mid-career teachers flattening out at between .1 and .2 standard deviations better than a novice teacher.³ More recent studies relax the functional form assumptions imposed in these early studies and find somewhat larger returns to high levels of teaching experience.⁴

However, this literature has generally ignored the possibility that the baseline effectiveness of a teacher and/or the gains to teaching experience might be specific to a particular classroom environment. In such a context, models that impose homogeneity of productivity across different classroom environments will return a weighted average of teacher productivity across the environments each teacher actually faced (weighted by the fraction of time spent in each environment). To the extent that teachers face different classroom contexts over their careers, models that impose homogeneity of returns to experience across different classroom environments may underestimate the gains to context-specific experience. Similarly, to the extent that teachers' classroom environ-

²For primary school estimates, see, for example, [Rockoff \(2004\)](#), [Hanushek et al. \(2005\)](#), [Clotfelter et al. \(2006\)](#), [Sass et al. \(2014\)](#), [Boyd et al. \(2008\)](#), [Jackson and Bruegmann \(2009\)](#), [Harris \(2009\)](#), [Harris and Sass \(2011\)](#), and [Jackson \(2013a\)](#). For secondary school estimates, see, for example, [Aaronson et al. \(2007\)](#), [Jackson \(2014\)](#), and [Mansfield \(ming\)](#). [Harris \(2009\)](#), by contrast, finds little evidence of returns to experience using high school data from Florida.

³e.g. [Rivkin et al. \(2005\)](#), [Clotfelter et al. \(2007\)](#).

⁴[Wiswall \(2013\)](#) and [Papay and Kraft \(2015\)](#).

ment remain somewhat stable during their career, such models may overestimate the returns to general experience.

A few papers, though, have addressed various aspects of the context-specificity of teacher productivity, mostly using elementary or middle school data. [Jackson \(2013a\)](#) shows that a substantial portion of the variation in teacher contributions to student achievement is specific to the school in which a teacher has taught. [Lockwood and McCaffrey \(2009\)](#) and [Aucejo \(2011\)](#) examine the degree to which teachers have comparative advantages at teaching relatively high versus low ability students, and find evidence that a small component of teaching productivity is specific to student ability level. Perhaps more closely related to our paper is work by [Ost \(2014\)](#) showing that teachers who always repeat elementary grade assignments improve 35% faster than teachers who never repeat grade assignments. Similarly, [Master et al. \(2012\)](#) show that the efficacy of a teacher teaching English-language learners (ELL) depends on his/her experience teaching the ELL population. The paper most closely related to ours is [Condie et al. \(2014\)](#), who also consider subjects as tasks. They demonstrate the existence of meaningful comparative advantages of elementary teachers at teaching English vs. math. These papers, however, focus either on context-specific experience or context-specific skill, rather than providing a unified treatment of both factors.

Given the applicability of our methodology to the broader worker-to-task assignment problem, our paper also contributes to a growing literature on task-specific human capital, brought to the forefront by [Gibbons and Waldman \(2004\)](#), which considers the possibility that a considerable portion of a worker’s human capital might be specific to the particular tasks the worker has performed at the jobs the worker has held.⁵ Part of the literature on task-specific human capital either has assumed that only the experience component of human capital is task-specific (e.g. [Gibbons and Waldman \(2004\)](#), [Clement et al. \(2007\)](#), and [DeAngelo and Owens \(2012\)](#)). Alternatively, [Polataev and Robinson \(2008\)](#) assume that the degree of task-specificity is common between the talent and experience components of human capital, while [Gathmann and Schoenberg \(2010\)](#) instrument to remove the influence of the task-specific talent component in order to focus on task-specific experience. [Yamaguchi \(2012\)](#) allows for both task-specific talent and gains to task-specific experience,

⁵See, for example, [Yamaguchi \(2012\)](#), [Clement et al. \(2007\)](#), [Polataev and Robinson \(2008\)](#), [Gathmann and Schoenberg \(2010\)](#), [DeAngelo and Owens \(2012\)](#).

but does not have productivity data, and thus must infer the size of each component indirectly from observed sequences of occupational choices.

To preview our results, we find that about a quarter to a third of the returns to years of experience are actually specific to the subject that the teacher taught. We find little evidence of returns to level-specific experience and no evidence of returns to subject-level experience. In agreement with the rest of the value-added literature, we find that the variation in fixed teaching skill is comparable in magnitude to the gains to experience. While 74% of the variance in permanent skill is general to all subjects and levels, we find a small but meaningful role for subject-specific (17%) and level-specific (9%) teaching talent: receiving a teacher whose subject-specific (level-specific) talent is one standard deviation above his/her average among all subjects (levels) he/she teaches increases a student's expected test score by .063 (.046) standard deviations.

We test for and fail to find convincing evidence of estimation biases driven by dynamic assignment responses to teacher-year or school-year shocks or unmodeled teacher-specific heterogeneity in gains from experience. Backcasting tests for bias from non-random student sorting to teachers suggest that, if anything, the significant gains to both general and subject-specific experience that we estimate may be understated. Split-sample forecast tests suggest that our estimates of teachers' combined general and task-specific talent have considerable out-of-sample predictive power, though admittedly slightly less than what a model with no bias or misspecification would imply. While similar split-sample forecast tests for teachers' estimated task-specific comparative advantages (more important for teacher assignment) are underpowered, they do not find evidence of any forecast bias in subject-specific talent estimates, though level-specific talent estimates do not seem to predict out-of-sample comparative advantages nearly as well.

Of course, the knowledge that a large fraction of the gains from experience are subject-specific may be of limited value to principals if most changes in course assignments are driven by necessity. For example, parental leave or teacher retirements may require principals to reassign teachers to unfamiliar subjects or tracks. Using our estimated experience profiles, we address this possibility by performing counterfactual simulations in which we reassign the teachers observed teaching in each school-field combination in the chosen year to the courses that were offered at their school at the time in order to maximize student performance, given posterior beliefs about the teachers' course-

specific talent as well as the four-dimensional stocks of general and context-specific experience that these teachers possessed at the beginning of the year.

Our simulations indicate that efficient use of task-specific experience and talent can, in principle, increase student achievement non-trivially: relative to random assignment of teachers to classrooms, the efficient allocation raises mean test scores by as much as .04 student-level standard deviations for school-field combinations with seven or more teachers. The degree to which principals' classroom assignments already effectively incorporate information about teacher comparative advantages is difficult to discern; however, under the strong assumption that the information about teachers' subject-specific and level-specific talent reflected in our statewide panel of 1997-2009 test scores is a superset of the information available to principals, our simulations suggest that efficient use of context-specific experience might increase mean test scores in larger high schools by as much as .02-.03 student-level standard deviations relative to the observed patterns of teachers' classroom assignments. Furthermore, since we hold the teaching load fixed for each teacher, these efficiency gains could potentially be reaped with near zero cost.⁶ These simulated gains are comparable in magnitude to the gains from subject-specialization in elementary school projected by [Condie et al. \(2014\)](#). We also show that they are comparable in magnitude to the gains administrators could expect to reap from a policy in which the least effective 10% of teachers are removed and replaced by average teachers.

The rest of the paper proceeds as follows. Section [3.2](#) presents the education production function whose parameters we estimate. Section [3.3](#) describes how comparisons of teachers with different course assignment histories can provide joint identification of both school-teacher-subject-level fixed effects and general, subject-specific, level-specific, and subject-level-specific experience profiles. Section [3.4](#) discusses the North Carolina administrative data and provides summary statistics illustrating the variation in teacher course assignments. Section [3.5](#) presents the parameter estimates from our main specifications. Section [3.6](#) presents tests for possible threats to our identifying assumptions and demonstrates the robustness of our results to alternative choices regarding sample

⁶Note that we cannot address the possibility that proposed reallocations would either detract from competing non-test score objectives or carry compensating differential costs (e.g. if teachers have strong preferences for teaching courses in their comparative disadvantages).

selection, variable definition, and model specification. Section 3.7 describes and presents results from the counterfactual simulations that gauge the test score gains that might be achievable through effective use of a teaching staff’s context-specific talent and experience. Section 3.8 concludes.

3.2 Model Specification

Because our goal is to determine the relative importance of context-specific teacher skill and experience for test score performance, we craft our specification of the achievement production function in a fashion that permits us to isolate the contribution of these components. Let Y_{ict} represent the standardized test score of student i in classroom c at time t . Let $r(i, c, t)$ denote the teacher that taught student i in classroom c at time t . Similarly, let $s(i, c, t)$ denote the school at which student i experienced classroom c at time t , let $j(i, c, t)$ denote the subject taught in student i ’s classroom c at time t , and let $l(i, c, t)$ denote the difficulty level or track associated with the classroom (Basic or Honors).⁷ Since North Carolina used different test forms for each subject in each year, we standardize each test score Y_{ict} so that the distribution of test scores in each subject-year combination has zero mean and unit variance.

By suppressing the dependence of s , r , j , and l on (i, c, t) , we can represent the production of test score performance compactly via:

$$Y_{ict} = X_{ict}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d^{gen}(exp_{rt}^{gen}) + d^j(exp_{rt}^j) + d^l(exp_{rt}^l) + d^{jl}(exp_{rt}^{jl}) + \epsilon_{ict} \quad (3.1)$$

Because we estimate the model at the classroom level, we aggregate this production function and focus our attention on the classroom-mean of test score performance, denoted Y_{ct} :

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d^{gen}(exp_{rt}^{gen}) + d^j(exp_{rt}^j) + d^l(exp_{rt}^l) + d^{jl}(exp_{rt}^{jl}) + \epsilon_{ct} \quad (3.2)$$

X_{ct} represents a vector of class-averages of student observable characteristics and middle school reading and math test scores, along with other classroom characteristics (e.g. class size $|I_c|$) and a full set of calendar year indicators. We allow the coefficients on X_{ct} , β_{jl} , to differ across subject-

⁷Section 3.4.2 describes how we assign courses to difficulty levels.

level combinations.⁸ This allows for the possibility that a high class average of 8th grade math scores might be a stronger predictor of class performance in Algebra 1 than in English 1. Similarly, classroom composition might matter more in a particular subject or level if more group work takes place in say, basic biology (e.g. labs) than in honors math. X_{ct} is included to control for non-random sorting of students to particular teachers within school-subject-level cells (discussed further in Section 3.3.2).⁹

δ_{sjl} represents inputs provided by the school-subject-level combination. The set of $\{\delta\}$ parameters will not only capture the contribution of any school-level inputs such as principal quality, neighborhood quality, or quality of the school facilities, they will also capture any variation in the quality of curricula or textbooks across subjects and levels within the school. δ_{sjl} will be estimated via a full set of school-subject-level fixed effects, $\hat{\delta}_{sjl}$. These fixed effects will capture the average residual achievement at each school-subject-level combination, after removing the part of achievement that can be predicted based on observable classroom characteristics. Importantly, in practice they will also reflect the contribution of average unobserved inputs of the students who sort into particular school-subject-level combinations. Thus, the school-subject-level design matrix also acts as a control function that absorbs school inputs as well as any potential sorting biases that might otherwise be created by students' endogenous choices of school, subject, and level.

μ_{srjl} captures the experience-invariant component of teacher r 's ability to increase student achievement in subject-level (j, l) at school s . μ_{srjl} will be estimated via a full set of school-teacher-subject-level fixed effects, $\hat{\mu}_{srjl}$. The average school-teacher-subject-level will be normalized to 0 for each school-subject-level in our baseline specification (see Section 3.3.2 for further discussion), so that $\hat{\mu}_{srjl}$ can be interpreted as the deviation of a particular teacher's performance in a particular subject-level combination from the mean (student-weighted) performance of all teachers that taught

⁸The coefficients on the calendar year indicators are restricted to be the same across all subject-levels to improve efficiency.

⁹Given that we include classroom averages of student inputs to better control for sorting on unobservable student characteristics (Altonji and Mansfield, 2014), aggregation of our outcome test scores to the classroom level is essentially without loss of generality. This is because the student-level observables are orthogonal to all the inputs of interest once class averages of these student observables have been conditioned on, since the inputs of interest display no within-class variation.

in the chosen teacher’s school-subject-level combination during the sample (e.g. how a particular honors biology teacher’s students performed relative to the honors biology students of his/her colleagues). This specification of the contribution of teacher quality allows the estimation of a fully non-parametric joint distribution of general teacher talent and subject-specific, level-specific, and even subject-level-specific permanent comparative advantages within and across teachers. Note that by including the identity of the school in the definition of the fixed effect, we are allowing each teacher’s mean contribution and comparative advantages for particular subjects and levels to be different at each school at which they teach (a teacher who teaches in two schools is essentially treated as two different teachers).

exp_{rt}^{gen} represents the total number of years of general teaching experience that teacher r possessed at the beginning of year t , defined as the number of previous years in which the teacher taught at least one course. Analogously, exp_{rt}^j , exp_{rt}^l , and exp_{rt}^{jl} represent previous years of experience teaching at least one course in the subject, level, and subject-level combination associated with classroom c , respectively. $d^{gen}(\cdot)$ is a function that captures the part of the gains from years of teacher experience that are portable (“general”) across all subjects and levels. The $d^j(\cdot)$, $d^l(\cdot)$, and $d^{jl}(\cdot)$ functions capture how additional years of subject-specific experience, level-specific experience, and subject-level-specific experience affect a teacher’s ability to increase student test scores. $d^{gen}(\cdot)$, $d^j(\cdot)$, $d^l(\cdot)$, and $d^{jl}(\cdot)$ are each flexibly parameterized using indicators for narrow ranges of experience.

Because the estimates from the “baseline” specification in (3.2) prove to be somewhat sensitive to choice of controls and the exact parametrization of the experience profiles, we also devote considerable attention to a less volatile “restricted” specification in which we constrain $\mu_{srjl} = \bar{\mu}_{sr} \forall (j, l)$ and (s, r) , allowing us to replace the school-teacher-subject-level fixed effects with school-teacher fixed effects only:

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \bar{\mu}_{sr} + d^{gen}(exp_{rt}^{gen}) + d^j(exp_{rt}^j) + d^l(exp_{rt}^l) + d^{jl}(exp_{rt}^{jl}) + \epsilon_{ct} \quad (3.3)$$

Finally, ϵ_{ct} represents the class average of an error component ϵ_{ict} that combines time-varying inputs not captured by the other components of the model. In particular, we model the class-level

error component as:

$$\epsilon_{ct} = \phi_{st} + \nu_{rt} + \zeta_{ct} + \frac{1}{|I_c|} \sum_{i \in c} e_{ict} \quad (3.4)$$

ϕ_{st} captures year-specific deviations in school inputs relative to the sample-wide average for the school-subject-level (e.g. due to school renovation). ν_{rt} represents year-specific deviations in a teacher's quality from what would be expected based on the teacher's time-invariant skill and context-specific experience (e.g. due to teacher illness). ζ_{ct} captures classroom level shocks, such as an uncontrollably disruptive student or the archetypal dog barking outside the classroom window on test day. Finally, e_{ict} represents the contributions of residual student-level inputs that are unpredictable based on observables as well as measurement error reflecting the extent to which the student's exam performance deviates from what the student could have expected to score, given his/her accumulated knowledge in the subject.

3.3 Identification

3.3.1 Identifying the Returns to General and Task-Specific Experience

Let $Exp = [Exp^{gen}, Exp^j, Exp^l, Exp^{jl}]$ represent the random vector of general and context-specific experience stocks for classroom teachers accumulated as of year t , of which each observed combination $[exp_{rt}^{gen}, exp_{rt}^j, exp_{rt}^l, exp_{rt}^{jl}]$ is a draw. Similarly, let M and D represent random vectors of school-teacher-subject-level and school-subject-level cell indicators, respectively. Each draw from M and D represents a row of the design matrices corresponding to the fixed effects capturing $\{\mu_{srjl}\}$ and $\{\delta_{sjl}\}$, respectively. Finally, let X represent the random vector of observed classroom characteristics, and let ϵ represent the random variable of which ϵ_{ct} is a draw. To identify the functions mapping experience stocks to productivity, $\{d^{gen}(*), d^j(*), d^l(*), d^{jl}(*), \}$, we assume that the following condition holds:

**Assumption 1: Conditional Mean Independence of
Time-Varying Unobserved Inputs and Teacher Experience**

$$E[\epsilon | Exp, M, D, X] = E[\epsilon | M, D, X] \quad (3.5)$$

Assumption 1 states that knowledge of the four-dimensional experience stock of the teacher does not provide further information about any unobserved inputs, conditional on observed classroom inputs and the identity of the school, teacher, subject, and level. Put another way, the timing of experience accumulation in each dimension of experience is assumed to be exogenous.

Recall from (3.4) that the regression error contains school-year, teacher-year, and classroom shocks (along with class-averages of individual-level unobserved inputs): $\epsilon_{ct} = \phi_{st} + \nu_{rt} + \zeta_{ct} + \frac{1}{|I_c|} \sum_{i \in c} e_{ict}$. Thus, there are a number of sources of possible threats to the validity of Assumption 1, each of which relates to the exact timing of changes in experience. For example, suppose that when a school is in decline, teacher turnover begins to increase, and the teachers that remain are forced to teach both new subjects and new difficulty levels more frequently. In this case, we may be more likely to observe zero subject-specific or level-specific experience when the contribution of time-varying school inputs ϕ_{st} is low. Since year-specific deviations in school quality from the sample-wide average are included in ϵ_{ct} , this scenario violates Assumption 1 and could potentially produce an overestimate of the returns to task-specific experience. Alternatively, suppose principals are reluctant to force a teacher to take on new subjects or levels when the teacher faces other short-term obstacles (such as illness or pregnancy). In that case, zero subject-specific or level-specific experience may be observed more frequently when the value of the teacher-year shock ν_{rt} is high. This scenario also violates Assumption 1, and might cause an underestimate of the returns to task-specific experience. Similarly, if teachers respond to a particularly unruly classroom by quitting teaching, or switching levels or subjects, we might underestimate the returns to experience (since those who survive to the next year of experience will have experienced above-average classroom shocks the previous year, thereby hiding the gains to the next year of experience). Finally, returns to experience could also be overestimated if students with superior unobserved inputs systematically sort into classes within subject-levels taught by teachers with high general or context-specific experience. We address the possibility of such violations of Assumption 1 in Section 3.6 and find little evidence of violations of sufficient magnitude to produce a substantial bias to any of our profiles.

Despite these concerns, however, note that Assumption 1 is still much weaker than the assumptions required to identify experience profiles in most of the literature, since it conditions on the combined identities of the school, teacher, subject and level. Essentially, the inclusion of school-teacher-subject-level fixed effects (μ_{srjl}) controls for any arbitrary selection of teachers into experience categories based on the fixed component of general or context-specific productivity. Conditioning on the identity of the teacher accounts for the possibility that better teachers persist long enough to gain more experience. Similarly, conditioning on the teacher-subject combination accounts for the possibility that the teachers allowed to gain more subject-specific experience in a particular subject are those with comparative advantages in teaching the subject, while conditioning on the teacher-level combination accounts for the possibility that persistence at teaching honors courses might signal a comparative advantage for teaching such courses.

Even if the timing of experience accumulation is conditionally independent of the error components, the simultaneous identification and estimation of each of the four experience profiles also requires considerable variation in the history of subject and level assignments across teachers. Such variation is necessary to satisfy the OLS rank condition and, more importantly, to produce sufficiently precise estimates. [A](#) illustrates how identification of the context-specific experience profile in each context dimension might be secured for our baseline model, and provides insight into the patterns of student performance in the data that inform estimates of the experience profile parameters.

The examples in [A](#) reveal that the experience profiles are fully identified from comparisons of different teachers' rates of performance growth (divergence/convergence of average student residuals) across years in which the same subject-level combination was taught. Because the average performance of each teacher in each school-subject-level combination is perfectly fit by the unrestricted school-teacher-subject-level and school-subject-level fixed effects, such cell averages provide no identifying variation for the experience profiles. Put another way, the inclusion of these fixed effects forces the identification of the experience profiles to be delivered exclusively from the path of productivity growth within school-teacher-subject-level combinations.

3.3.2 Identification of the General and Context-Specific Components of Fixed Teaching Skill

Identifying fixed or pre-determined general and context-specific teaching skill is more difficult. In particular, there is a fundamental identification problem that our model cannot overcome: we cannot distinguish average teaching quality in a particular school-subject-level from school or unobserved student inputs that vary across school-subject-level cells. For example, suppose a school's students score 0.1 student-level standard deviations higher in Biology than in Chemistry. In the absence of restrictions on the distribution of subject-specific teacher skill, we cannot tell whether all the teachers at the school are particularly effective at teaching Biology relative to Chemistry, or if instead the Biology textbook is superior to the Chemistry textbook (or many of the student's parents are biologists). To address this issue, we perform a sensitivity analysis in which we introduce two polar opposite assumptions and one moderate assumption for apportioning the between school-subject-level achievement variation between teachers and other inputs. We decompose the variance in teacher time-invariant productivity into general, subject-specific, level-specific, and subject-level-specific components under each assumption.

The first extreme assumption is that average teacher effectiveness is uniform across all levels, subjects, and schools:

Assumption 2A: Uniform Average Teacher Quality Across Contexts

$$\frac{1}{|I_{sjl}|} \sum_{(i,c,t) \in (s,j,l)} \hat{\mu}_{srjl} = k \text{ for some constant } k, \forall (s,j,l) \in \mathcal{S}\mathcal{J}\mathcal{L} \quad (3.6)$$

where $|I_{sjl}|$ is the number of students observed taking subject j in level l at school s and $\mathcal{S}\mathcal{J}\mathcal{L}$ is the set of all school-subject-level combinations. This assumption will hold (with a sufficiently large pools of teachers) if the relatively more effective teachers do not sort into particular schools, subjects, or levels. Assumption 2A implies that all the variation in average residual student performance across subjects, levels, and schools (after removing the part that is predictable based on classroom observables) can be attributed to either school inputs or unobserved student inputs. Assumption 2A can be imposed on the model by estimating a full set of school-subject-level fixed effects ($\hat{\delta}_{sjl}$), and normalizing the student-weighted average teacher-school-subject-level fixed effect to be zero at each

school-subject-level: $\frac{1}{|I_{sjl}|} \sum_{(i,c,t) \in (s,j,l)} \hat{\mu}_{srjl} = 0$ (note that the common mean k does not contribute to variance estimates). Under Assumption 2A, if the school average performance difference between Biology and Chemistry is 0.1 standard deviations then a teacher whose Biology students perform 0.1 standard deviations better than her Chemistry students will be assumed to be equally effective at teaching both Biology and Chemistry.

A second intermediate assumption assumes that between-school variation in residual test scores is fully attributable to school quality and student sorting, but that the variation in residual performance that is within-schools but across subject-level combinations is fully attributable to differences in average teacher quality across these combinations:

Assumption 2B: Uniform Teacher Quality Across Schools, Uniform Student/School Quality Across Subjects and Levels

$$\begin{aligned} \delta_{sjl} &= \bar{\delta}_s \quad \forall (s, j, l) \in \mathcal{S}\mathcal{J}\mathcal{L} \\ \frac{1}{|I_s|} \sum_{(i,c,t) \in s} \mu_{srjl} &= k \text{ for some constant } k, \quad \forall s \in \mathcal{S} \end{aligned} \tag{3.7}$$

Estimates from such a model are useful for a principal who needs to determine classroom assignments for his/her existing stock of teachers. The principal will only require the decomposition of the within-school variance in time-invariant teacher productivity, and may believe that school inputs are divided relatively equally across subjects and levels. Assumption 2B is implemented by replacing the school-subject-level effects with school fixed effects only, and restricting the average school-teacher-subject-level effect to be 0 at each school.

Finally, the opposite extreme approach is to assume that all the variation in average residual student performance across subjects, levels, and schools can be attributed to differences in average teacher quality:

Assumption 2C: Uniform School and Unobserved Student Quality Across Contexts

$$\delta_{sjl} = k \text{ for some constant } k, \quad \forall (s, j, l) \in \mathcal{S}\mathcal{J}\mathcal{L} \tag{3.8}$$

Assumption 2C will hold if students sort into high schools, subjects, and levels based only on observable characteristics and past performance, and all high schools and subject-level combinations within high schools provide the same contribution to student achievement. Assumption 2C can be imposed on the model by excluding school-subject-level fixed effects entirely ($\hat{\delta}_{sjl} = 0 \forall (s, j, l)$), and matching the between school-subject-level residual variation using a full set of teacher-school-subject-level fixed effects (without any normalizations). Under Assumption 2C, a teacher whose Biology students perform 0.1 standard deviations better than her Chemistry students will be assumed to be 0.1 standard deviations more effective at teaching Biology than Chemistry if the school average performance difference between Biology and Chemistry is 0.1 standard deviations. In other words, even though the teacher is at the mean of the performance distribution in both subjects, the comparison set of Biology teachers is assumed to be 0.1 standard deviations superior on average to the comparison set of Chemistry teachers.

While Assumptions 2A-2C allow us to separate school inputs from teacher inputs, identification of $\{\mu_{srjl}\}$ also requires that other unobserved inputs are not correlated with the observation of a particular teacher in a particular subject-level combination. As before, M and D represent the random vectors of school-teacher-subject-level and school-subject-level cell indicators, while Exp represents the random vector of teacher experience stocks and X represents the random vector of observed classroom characteristics. Similarly, let S represent the random vector of school indicators (draws of which would represent a row of a design matrix for a set of school fixed effects). Then assumptions 3A-3C capture this additional condition for each of the three cases considered:

**Assumption 3A-3C: Conditional Mean Independence of
Students' Unobserved Inputs and Teacher Identity**

$$\begin{aligned}
3A : \quad & E[\epsilon|M, Exp, X] = E[\epsilon|D, Exp, X] \\
3B : \quad & E[\epsilon|M, Exp, X] = E[\epsilon|S, Exp, X] \\
3C : \quad & E[\epsilon|M, Exp, X] = E[\epsilon|Exp, X]
\end{aligned} \tag{3.9}$$

Assumption 3A states that the identity of the teacher does not provide further information about any unobserved inputs, conditional on the identities of the school, subject, and track, along with the levels of general and context-specific experience of the teacher and the observable classroom characteristics. Note that by conditioning on all four dimensions of teacher experience, we remove the concern that a teacher will be perceived to have greater general skill because he/she has more general experience, or that a teacher will be perceived to have a comparative advantage at teaching in a particular context because many of the test-score observations from that context are accompanied by considerable context-specific experience. Assumption 3B is much stronger, since it conditions on the school rather than the school-subject-level, while Assumption 3C, which conditions only on teacher experience stocks and observed classroom inputs, is the strongest assumption of all.

There remain several potential threats to the validity of the fixed effect estimates even in the case of Assumption 3A. Suppose, for example, that a particular teacher is assigned to a room with broken air conditioning each time the teacher teaches honors physics, but is assigned to functioning rooms whenever the teacher teaches honors chemistry. In this case, conditioning on context-specific experience will not remove the correlation between the classroom-level error component ζ_{ct} and the indicator for the school-teacher-subject-level combination associated with the chosen teacher teaching honors physics. Similarly, a teacher who happens to be assigned to basic English 1 classes during the years her kids are young (when she has little time to prepare for class) might exhibit a correlation between the unobserved teacher-year shock ν_{rt} and the indicator for the school-teacher-subject-level combination associated her basic English 1 course.

Perhaps the most serious concern stems from the possibility that unobservably superior students are able to disproportionately select a particular teacher.¹⁰ This possibility is somewhat less likely at the high school level, since class assignments are frequently generated by scheduling algorithms (given students' subject-level choices), making it difficult for students to select into particular classrooms within a subject-level. We rely on classroom averages of student covariates to absorb

¹⁰Rothstein (2010) documents non-random student sorting into particular classrooms within North Carolina elementary schools. However, Kinsler (2012) retests the same data, accounting for small sample sizes, and fails to reject such non-random sorting.

any within-subject-level sorting based on student inputs. [Altonji and Mansfield \(2014\)](#) show that classroom averages of observable characteristics can in theory absorb all the bias from sorting on both observables *and* unobservables, if the set of observables is diverse enough to span the set of classroom amenities that are driving students to sort. Furthermore, [Mansfield \(ming\) and Jackson \(2014\)](#), using the same NCERDC dataset we employ here, find little evidence of remaining student sorting after controlling for track. In [Section 3.6.2](#), though, we investigate further the possibility that sorting of students to teachers could bias our estimated production function.

[B](#) provides a concrete example that illustrates the kinds of moments in the data that identify time-invariant teaching skill. The example in [B](#) reveals that each school-teacher-subject-level fixed effect $\hat{\mu}_{srjl}$ will be estimated using only a single teacher’s performance during the few years in which they taught the subject-level associated with the fixed effect. As such, sampling error for any given fixed effect estimate $\hat{\mu}_{srjl}$ will not converge to zero even with the fairly long panel we employ. Consequently, we do not focus on individual $\hat{\mu}_{srjl}$ estimates, but instead seek to characterize the joint distribution of the components of time-invariant teaching skill. Specifically, we decompose the variance in performance across teachers and contexts into components attributable to general teaching talent, subject-specific talent, level-specific talent, and subject-level-specific talent.

To see how this may be done, note first that we can rewrite the true value of teacher r ’s context-specific productivity μ_{srjl} via:

$$\mu_{srjl} = \bar{\mu}_{sr} + (\mu_{srjl} - \bar{\mu}_{sr}) \quad (3.10)$$

The first component in [\(3.10\)](#) can be interpreted as the contribution of teacher talent that may be school-specific, but is general or portable across tasks (subject-level combinations) within the school. We will refer to $Var(\bar{\mu}_{sr})$ as the variance in general teaching talent. The second component contains the teacher’s persistent subject-level-specific deviation in quality from the teacher’s average across all subject-level combinations. This can be interpreted as the teacher’s comparative advantage or disadvantage at teaching subject-level combination (j, l) . This second component can then be decomposed into three further components:

$$(\mu_{srjl} - \bar{\mu}_{sr}) \equiv \tilde{\mu}_{srjl} = \bar{\tilde{\mu}}_{srj} + \bar{\tilde{\mu}}_{srl} + (\tilde{\mu}_{srjl} - \bar{\tilde{\mu}}_{srj} - \bar{\tilde{\mu}}_{srl}) \quad (3.11)$$

The first component of (3.11) can be interpreted as the part of the teacher’s comparative advantage at subject-level combination (j, l) that is portable across levels but not subjects. We will refer to $Var(\bar{\mu}_{srj})$ as the variance in subject-specific teaching talent. The second component of (3.11) can be interpreted as the part of the teacher’s comparative advantage at subject-level combination (j, l) that is portable across subjects but not levels. We will refer to $Var(\bar{\mu}_{srl})$ as the variance in level-specific teaching talent. The third component of (3.11) is the part of a teacher’s comparative advantage at (j, l) that is not portable across levels or subjects, and thus could not have been predicted based on the sum of the teacher’s subject-specific skill and the teacher’s level-specific skill. We will refer to $Var(\tilde{\mu}_{srjl} - \bar{\mu}_{srj} - \bar{\mu}_{srl})$ as the variance in subject-level-specific teaching skill.

Note that we do not observe the true variance of school-teacher-subject-level effects, $Var(\mu_{srjl})$, but rather the sample variance, which contains sampling error: $Var(\hat{\mu}_{srjl})$. To recover the true latent variance decomposition, we follow the method of Aaronson et al. (2007) and Mansfield (ming). C describes this sampling error correction in detail.

Because we can only estimate a value of $\hat{\mu}_{srjl}$ for those school-teacher-subject-level combinations that we actually observe in the data, the variance in subject-specific and level-specific skill that we estimate will represent the variance among the range of subject and level combinations that principals actually assign.

While we are likely to underestimate the variance in subject-specific (or level-specific) talent across the full range of possible subjects (or levels), the estimates we do obtain are more relevant and interesting to principals and administrators. Much of the missing variance stems from variation in the strength of teacher’s comparative disadvantages among classroom assignments that are never seriously considered by principals (i.e. variation among English teachers in their ability to teach physics). Rather, the choice principals generally face is between hiring a new teacher to teach exactly the courses taught by an exiting teacher and hiring a new teacher to teach different courses while rotating existing teachers who are certified in the chosen field to new subjects or levels within that field (for example, rewarding stayers by letting them teach the honors class that was vacated by the exiting teacher).¹¹ Given the limited support for the distribution of comparative advantages

¹¹Teacher certification in North Carolina, as in most states, is at the level of the field (math, science, history, etc.) rather than the subject (Biology, Chemistry, Physics), and is not specific to a level of difficulty (special education

that underlies our estimates, in our simulations in Section 3.7 we only reallocate teachers across classrooms within fields.

3.4 Data

3.4.1 Overview

The decomposition of worker productivity developed in Sections 3.2 and 3.3 requires that the data 1) contain signals of worker output in each task, 2) allow the construction of accurate measures of general and task-specific experience, and 3) exhibit considerable worker rotation among tasks. We employ administrative data provided by the North Carolina Education Research Data Center (NCERDC) that satisfies each of these three conditions for the context of high school teaching.

3.4.2 Task-Specific Output and Sample Restrictions

The NCERDC data consists of standardized test scores for the universe of public high school students in North Carolina from 1997 - 2009 in eleven subjects and two course difficulty levels.¹²

During the sample period, North Carolina provided a standardized curriculum in each subject and assessed achievement via statewide end-of-course tests.¹³ The eleven subjects, which can be

excepted).

¹² The student-level End-of-Course test data provide a set of four difficulty level categories (honors, AP, college placement, and other) that do not perfectly match the difficulty level categories provided with the beginning-of-year classroom data (Special Education, Remedial, Basic, Applied/Technical, Honors, Cooperative Education, Advanced Placement, International Baccalaureate, and Non-Classroom), which contain the correct teacher ID (on which the level-specific experience stocks are based). In order to minimize the probability that the relevant level-specific experience of the teacher is mismeasured, we drop student observations coming from classes labeled as Special Education, Cooperative Education, and Non-Classroom. We classify Remedial, Basic, and Applied/Technical classes as “basic” and Advanced Placement, International Baccalaureate, and Honors as “honors”. In the rare cases where schools offer distinct Advanced Placement and Honors courses in the same tested subject we drop observations from classrooms where the teacher’s relevant level-specific experience depends on whether these two difficulty levels are combined during the construction of level-specific experience stocks.

¹³Note that these tests are subject-specific but not level-specific.

grouped into four fields based on common certification requirements, are as follows: *Math*: Algebra 1, Algebra 2, Geometry; *Science*: Biology, Chemistry, Physical Science, Physics; *Social Studies*: Econ/Law/Politics, Civics and Economics, U.S. History; *English*: English 1.¹⁴ Because statewide achievement tests were administered immediately at the conclusion of each year-long course, and the subjects are (largely) distinct from one another, average student performance in each course represents a signal (albeit a noisy, possibly biased one) of the task-specific output of the teacher.

In our framework, accurately distilling the signal of a teacher’s task-specific productivity from student sorting requires rich data on student inputs. Fortunately, the NCERDC data contain information about a variety of current student inputs (or proxies for such inputs), as well as past student inputs.¹⁶ In addition, we also include in X_{ct} the number of classes and number of distinct courses taught contemporaneously by the student’s teacher in order to capture teacher workload, and include indicators for whether the student’s teacher taught the current subject, level, and subject-level (and whether he/she taught at all) in the previous year to capture depreciation of human capital. We also include in X_{ct} a full set of dummies for the calendar year t in which the test was taken, as emphasized by Papay and Kraft (2015).

Properly measuring teacher contributions to achievement also requires that each student test score

¹⁴Testing began for Physics, Geometry, Chemistry, Physical Science, and Algebra 2 in 1999. In addition, Econ/Law/Politics was discontinued in 2004 and replaced by Civics and Economics in 2006. U.S. History was not tested between 2004 and 2005.

¹⁵In principle, one might worry that differences in teacher performance may be reflecting the extent to which teachers adhere to the state curriculum rather than differences in ability to foster learning. Fortunately, several features of the North Carolina context mitigate such concerns. First, in recent years No Child Left Behind legislation has put pressure on principals to ensure that teachers teach the standard curriculum, since schools that fail to meet state standards are subject to sanctions and possible closure. Second, the North Carolina end-of-course exam scores we use as outcome measures must comprise 25% of the student’s year-end grade in a given subject, so that parents are likely to complain about teachers that ignore the standard curriculum. Finally, during the sample period, teacher bonuses of up to \$1,500 were linked to average test scores of the students in the school at which they teach.

¹⁶Observable current student inputs include indicators for parental education, race, gender, gifted status, current grade, and current Limited English Proficiency status. Observable past student inputs include the student’s 7th and 8th grade math and reading scores (though to reduce the influence of missing data we only include 7th grade test scores as a robustness check in Section 3.6.4).

observation be matched to the teacher who taught the class in which the student’s test score was generated. We utilize the fuzzy matching algorithm developed by Mansfield (ming), which exploits the fact that classroom-average demographics can be constructed and compared for both the test-score-level data and the classroom-level data (which contains the valid teacher ID).¹⁷

Our original dataset consists of 8,407,382 test scores from 460,792 classrooms, 28,347 teachers, and 1,307 high schools. We drop from the sample 2,878,254 test score observations for which we cannot match a teacher and 794,541 for which we cannot verify a difficulty level (discussed in footnote (12) above).

In addition, recall that the second key data requirement is that measures of both general and task-specific experience can be accurately constructed. The NCERDC data contain all classroom assignments (subject and level) for each teacher for the years 1995-2009, even in non-tested subjects. However, complete histories of classroom assignments, necessary to construct subject-specific, level-specific, and subject-level-specific experience, can only be assembled for teachers who began teaching after the data collection commences in 1995 (as indicated by an entry level paycode). Because our identification strategy relies on observing each teacher’s full history of subject- and level-specific experience at each point in time, we drop an additional 2,364,544 test scores associated with teachers for whom we cannot properly construct context-specific experience stocks. Note that we cannot distinguish novice teachers from teachers who previously taught outside of North Carolina unless such transferring teachers are given partial credit for their prior experience (and thus would not have an entry level paycode). Nonetheless, the problem of accurately constructing stocks of context-specific experience would be considerably more severe in contexts where data exist only for a single school district (even a large one).

After several other sample restrictions, our final sample consists of 1,126,300 test scores aggregated to 61,993 classroom-level observations, from 8,750 teachers, and 596 high schools.¹⁸ Basic summary

¹⁷See Mansfield (ming) for a full description of the algorithm and summary statistics regarding its efficacy.

¹⁸We restrict the sample in several additional ways. First, we drop 21,915 scores from classes with fewer than 5 students (since these are likely to represent data entry errors). Given our focus on high schools, we also drop 263,893 test scores from students in grades 6-8. We also drop test scores with invalid or outlier values, as well as all scores from 1997 and from Physical Science in 1999 due to concerns about data quality (270,395 scores). Since past test scores are critical for controlling for student sorting, we also drop 685,116 observations for students with missing 8th grade

statistics comparing the original and final samples are presented in Table 3.1.

3.4.3 Generating the Experience Profile

For the baseline specification we construct flexible experience profiles by creating indicators for eight experience categories: 0 years of experience, 1 year, 2 years, 3 years, 4 years, 5-6 years, 7-10 years, and 11 or more years of experience. In our featured specifications, experience is measured as the number of prior years in which at least one classroom was taught in the relevant context for the chosen experience dimension. We posit that teaching a second classroom in the same year, when there is no opportunity to alter the lesson plan or assignments, is likely to provide negligible experience value relative to teaching a classroom in a different year. However, as a robustness check we also present results from specifications in which experience is measured using the total number of classrooms taught prior to the year of the observation.

We also assume that teachers' general and context-specific experience is fully portable across schools. This assumption is partly driven by the existence of a statewide testing regime that is tied to students' course grades, so that curriculum differences between schools should be minimal. Further, since only 16% of our classrooms are taught by a teacher who in his/her second (or greater) school, we also have limited variation with which to test this assumption. However, in Section 3.6.4 we examine the sensitivity of our results to this assumption by estimating our general and context-specific experience profiles for only the subset of classrooms featuring teachers in their initial schools.

To capture depreciation in teachers' "experience capital", we include in X_{ct} a set of indicators for whether the teacher of the classroom taught the subject, level, and subject-level in the previous year, as well as an indicator for whether the teacher taught at all in the previous year. We also include a second set of analogous indicators for whether the teacher of the classroom taught in the

math or reading test scores. Finally, identification of experience cell fixed effects (estimated in our "full specification" discussed in section 3.6.4 below) requires that four-dimensional experience cells and school-teacher-subject-level cells form a connected graph, with the experience cells as vertices and school-teacher-subject-level cells as edges (or vice versa). We drop 2,424 test scores that are associated with school-teacher-subject-level combinations not contained within the largest connected component of the graph.

relevant contexts two years prior.

Finally, to account for possible decreases in teacher effort prior to an assignment change (explained further in Section 3.6.1) we also include four indicators that equal one if the observation is from a classroom that represents the teacher’s last year teaching the school-subject combination, the school-level combination, the school-subject-level combination, and at the school in any classroom, respectively.

3.4.4 The Frequency of Teacher Assignment Rotations

The third data requirement for our decomposition is that we observe considerable worker rotation across tasks. Table 3.2 depicts teacher rotation across subjects in our final sample. The top (bottom) entry in each cell (i, j) represents the number (fraction) of teachers in our sample who ever taught in subject i that also taught in subject j . The table reveals that there is considerable rotation across subjects, though the vast majority of rotations occur within fields. This reflects the fact that certification is field-specific. Teacher rotation across levels is also substantial. The vast majority (87%) of teachers who ever teach an honors class also teach at least one basic class during their career. The converse is not true; only 43% of teachers observed teaching at least one basic class are also observed teaching an honors class at some point during their careers. This finding partly reflects the fact that there tend to be more basic courses than honors courses to staff at most schools, but is also driven by a substantial fraction of schools that do not track their classes (so that all classrooms at the school are coded as being taught at the basic level).

Table 3.3 displays the pattern of rotation across subject-level combinations for teachers in the field of mathematics. The table illustrates that teachers do not merely teach either multiple levels of the same subject or multiple subjects at the same level, but rather are frequently observed teaching at the basic level in one subject and at the honors level in a different subject. It is this variation that allows us to distinguish the returns to subject-level-specific experience from the returns to subject-specific and level-specific experience, respectively. Taken together, these tables demonstrate that rotating across multiple subjects, levels, and subject-levels during one’s career is the norm, rather than the exception.

As the example in [A](#) makes clear, identification of all four dimensions of experience relies on teachers continuing to introduce new subjects and levels into their repertoire after their career is already underway, as well as taking single year or multi-year breaks from teaching particular subjects before returning to them later.¹⁹

Figure [3.1](#) shows, for each level of general experience, the fraction of teachers who teach a subject, level, or subject-level for either the first time ([3.1a](#)) or last time ([3.1b](#)), as well as the fraction of teachers who leave teaching in North Carolina altogether (“General” in [3.1b](#)). Figure [3.1](#) reveals that introducing new courses is quite common even in mid-career: 19% of teachers with seven prior years of experience teach a new subject for the first time in their eighth year, while 11% teach a new level and 29% teach a new subject-level combination.

Gap years in which teachers fail to teach (and then return to) a particular subject or level are also quite common. 22.5% of unique teacher-subject-level combinations exhibit one or more gap years at some point during our sample, while 19.9% and 14.1% of teacher-subject and teacher-level combinations exhibit at least one gap year. By contrast, 10.2% of observed teachers leave public school teaching entirely for at least a year before returning. These statistics reveal that there is more variation available to identify the returns to subject- or level-specific experience than there is to identify gains that are portable across all contexts ([Wiswall \(2013\)](#), [Papay and Kraft \(2015\)](#)).

Finally, Figures [3.2a](#) and [3.2b](#) display the distributions of subject-specific, level-specific, and subject-level specific experience for classrooms taught by second and third year teachers in our final sample. The data underlying these figures are presented in Appendix Table [F.2](#). About 71% of classrooms taught by 2nd year teachers are in subject-level combinations that these teachers taught in their first years, while 55% of classrooms taught by 3rd year teachers are in subject-level combinations that these teachers taught in both of their first two years.

¹⁹To see this clearly, note that if every teacher taught the same exact subject/level combinations each year for their entire career, level-specific, subject-specific, and subject-level-specific experience would all increment by one every year, and would thus be perfectly collinear with general experience. By contrast, the relative within-teacher performance among multiple courses taught simultaneously provides an important source of variation in identifying the variance in permanent task-specific talent.

3.4.5 Estimation and Calculation of Standard Errors

We estimate the model at the classroom level via weighted OLS by exploiting the sparsity of the design matrices for the school-subject-level and school-teacher-subject-level fixed effects. Weights for each classroom observation are proportional to the number of students in the classroom, so that the variances in teacher productivity presented below capture the variation in teacher contributions across student-course combinations. Cluster-robust standard errors are calculated for each parameter. We cluster at the teacher level in order to accommodate the possibility of autocorrelated teacher-year shocks.

3.5 Results

3.5.1 Variation in the General and Context-Specific Components of Time-Invariant Teacher Productivity

Table 3.4 contains the results of the decomposition of the variance in time-invariant teacher productivity (“talent”) into general, subject-specific, level-specific, and subject-level-specific components using the baseline specification (3.2). The first column displays the decomposition obtained from imposing Assumption 2A, in which all between school-subject-level variation in student performance is attributed to differences in school and unobserved student inputs. The row labeled “School-Teacher-Subject-Level Combos” provides the total estimated variance (and corresponding standard deviation) in time-invariant teacher contributions to test scores across randomly sampled student-course combinations, which combines all four components of time-invariant teacher productivity. A one standard deviation increase in combined permanent teaching effectiveness is associated with a .154 standard deviation increase in expected student performance. 74% of this variance in permanent teacher quality can be attributed to general teacher talent that is portable across all subject-level combinations (see the row labeled “General Talent”). A student assigned to a teacher whose average effectiveness across the subject-level combinations he/she teaches is one standard deviation above the school average can expect a .132 standard deviation increase in test score performance relative to being assigned the average teacher at the school in the absence of

knowledge about the chosen teacher’s experience or level-specific and subject-specific skill.

Subject-specific skill and level-specific skill make up about 17% and 9%, respectively, of the total variance in permanent teaching effectiveness across randomly chosen student-course combinations (tests). Receiving a teacher whose subject-specific skill in the selected subject is one standard deviation above the teacher’s subjectwide average increases expected student achievement by about .063 test score standard deviations. Note that this is still enough to move a student who would have otherwise scored at the 50th percentile to the 53rd percentile statewide. Getting a teacher whose level-specific skill is one standard deviation above his/her levelwide average increases expected performance by .045 test score standard deviations, enough to move a student from the 50th to the 52nd percentile.

Finally, the subject-specific, level-specific, and general components of time-invariant teacher productivity combine to explain nearly the full variance in time-invariant teacher productivity across classroom contexts. Subject-level-specific talent does not seem to exist. In other words, a teacher’s permanent talent for teaching, say, honors biology, can be fully explained by the teacher’s general teaching talent across subjects and levels, combined with the teacher’s talent for teaching honors-level courses and the teacher’s talent for teaching biology courses, respectively.

Columns 3 and 4 of Table 3.4 display the alternative decomposition of permanent teacher skill that comes from imposing Assumption 2B, in which all the variation in average student performance across subject-level combinations within schools is also attributed to differences in average teacher quality. Not surprisingly, this increases each of the variance components substantially. Note, though, that the fractions of variance in teacher productivity explained by each component stay roughly similar to what they were under Assumption 2A. Under Assumption 2B, a one standard deviation increase in general teacher talent is associated with a .192 increase in average student performance, while a one standard deviation increase in subject-specific (level-specific) teacher talent is associated with a .077 (.058) increase in expected student performance relative to a subject (level) in which the teacher has no comparative advantage or disadvantage. Subject-level-specific talent does not appear to exist under Assumption 2B either. These results are roughly in line with those of Mansfield (ming).

The results under Assumption 2C (Columns 5 and 6) assign all the between school-subject-level variation in student performance to differences in teacher inputs rather than school or student inputs. They provide an upper bound estimate of the standard deviation in general teacher talent of .225 test score standard deviations.

Overall, we conclude that most of the time-invariant variation in teacher productivity is portable across all subjects and levels, but that there is a non-negligible achievement gain from being taught by a teacher who is relatively well-matched to the level and particularly the subject associated with the classroom.

3.5.2 General and Context-Specific Experience Profiles

Table 3.5 presents the estimated experience profiles for each type of experience from the baseline specification (3.2).²⁰ Panel A of Figure 3.3 displays these experience profiles graphically. Column 1 of Table 3.5 contains estimates of the part of the returns to teaching experience that are portable to all subject-level combinations, while Columns 2-4 contain estimates of the part of the returns to teaching experience that are subject-, level-, and subject-level-specific, respectively. There are considerable gains from the first two years of general experience, such that teachers teaching in their third year can expect to improve student performance by .085 test score standard deviations more than a novice teacher, even if they are teaching at a new level in a new subject. These gains grow to .113 by 7 years of experience, but seem to plateau thereafter. However, the results become quite noisy for higher levels of experience; since we must observe the entire history of teacher assignments, only the cohorts of new teachers from the late 1990's are observed at the higher levels of experience in our sample.

Row 1 of Column 2 indicates that teaching a subject for the second time increases the teacher's expected performance by .014 test-score standard deviations within that subject, relative to the first attempt. An additional year of subject-specific experience increases performance by an additional .019 standard deviations, while a third year of subject-experience adds an additional .016 standard

²⁰The coefficients on our controls for teacher workload and depreciation of experience capital for this specification are presented in Appendix Table F.3.

deviations. Gains seem to slow beyond the third year of subject experience. Overall, teachers with more than 7 years of subject-specific experience are between .046 and .067 student level standard deviations more effective than teachers with the same total years of general teaching experience but who are teaching the subject for the first time.

The results in Column 2 suggest that part of the returns to experience generally estimated in the literature are actually specific to the subject taught. Since teachers frequently reteach the same subject many times, subject-specific experience and overall (general) years of experience are highly correlated. Thus, when returns to subject-specific experience are not separated from returns to general experience, the returns to subject-specific experience will generally be reflected in larger estimated returns to general experience.

Columns 3 and 4, by contrast, show that the returns to level-specific and subject-level-specific experience seem to be virtually non-existent, once years of subject-specific and general experience have been taken into account. In fact, the returns to subject-level-specific experience seem to be negative. Note that such negative returns are not implausible in principle: teaching the exact same course again and again could cause teachers to lose enthusiasm or to stop updating course materials (even as the state curriculum drifts slightly).

That said, this negative profile might also be spurious if it is merely the product of overfitting; while including a full set of school-teacher-subject-level fixed effects removes potential bias from teachers systematically repeating the courses at which they are relatively effective more frequently, it also considerably limits the remaining variation in experience stocks that can be used to identify gains from experience.²¹ Given that subject-specific and subject-level-specific experience are very highly correlated, OLS may be able to reduce squared residuals more by fitting sampling error than by fitting true productivity gains.

To address concerns about overfitting, we turn attention to our “restricted” specification (3.3) that replaces the school-teacher-subject-level effects μ_{srjl} with school-teacher fixed effects only.²²

²¹Specifically, the inclusion of these fixed effects implies that only relative growth rates in performance within a school-teacher-subject-level cell provide identifying variation.

²²Consistent estimation of experience profiles in the restricted specification requires that teachers do not systematically gain more general or context-specific experience in the subjects or levels in which they have experience-invariant

Because the results from the restricted specification have proven to be more robust to alternative sample restrictions and the inclusion of additional controls, we focus primarily on experience profiles that maintain these restrictions for the remainder of the paper.

Table 3.6 displays the estimated general and context-specific experience profiles for the restricted specification (with Panel B of Figure 3.3 providing a graphical depiction). The results for general and level experience are essentially unaffected by the restrictions, but the negative effects of subject-level-specific experience disappear, while the gains to subject-specific experience are somewhat diminished. Specifically, a teacher with two (four) prior years of subject-specific experience could be expected to increase achievement by .023 (.041) test score standard deviations relative to the teacher's expected performance when teaching the subject for the first time (holding the other experience components fixed).

Column 5 in Table 3.6 sums across the first four columns to provide the returns to experience for a teacher who never changes the subject-level he/she teaches. After two (four) years, such a teacher is predicted to perform .118 (.138) standard deviations better than a novice teacher. Since many teachers teach the same subject-level every year (perhaps in addition to other courses), this sum is particularly well identified. Most of the sampling error in the estimates comes from decomposing this sum into the four experience components.

Given the failure to observe meaningful level-specific and subject-level-specific experience effects, the first two columns of Table 3.7 display results from a yet more parsimonious specification in which the level-specific and subject-level specific experience profiles are constrained to be zero everywhere. The basic pattern of results for total and subject-specific experience exhibit little change; there are still meaningful gains from the first several years of both total experience and subject-specific experience. Imposing these further restrictions increases the precision of the estimates considerably, however, so that experienced teachers are statistically significantly more effective than novice teachers for all categories of general experience and for all but the highest experience category of subject-specific experience.

comparative advantages. However, given that the previous sub-section revealed relatively small variances in subject- and level-specific permanent talent, even substantially elevated rates of re-assignment of teachers to their more effective subjects and levels would produce minimal bias.

The fourth column of Table 3.7 presents estimates from the standard specification in the literature, in which only a single “general” experience term enters the production function. This standard experience profile, which is driven by both general and specific returns, matches fairly closely those found in the literature.

Overall, the relative magnitudes of the coefficients for the different dimensions of experience parallel the results for context-specific talent presented in Section 3.5.1: a large role for the general component, with a moderate role for the subject-specific component and small-to-nonexistent roles for the level-specific and subject-level-specific components.

3.6 Tests of Identifying Assumptions and Robustness Checks

3.6.1 Testing for Dynamic Classroom Assignment Responses to Unobserved Shocks

Assumption 1, which is necessary for consistent estimates of experience profiles, will be violated if particular experience profiles are more likely to be observed during years in which either teachers or their schools are experiencing positive or negative year-specific deviations in productivity relative to what could be predicted given their full sample performance and teachers’ observed levels of each dimension of experience.

There are a variety of scenarios that could bring about such a correlation. Some involve endogenous allocation responses to idiosyncratic shocks, and may not exhibit any pre-trend. For example, a teacher who is less effective while pregnant may quit teaching after the baby arrives. Scenarios such as these would imply that the set of teachers who make it to the next year of teaching (or perhaps teaching in a particular context) are those whose teacher-year (or perhaps classroom) shocks were not too negative. Thus, the expected change in the teacher-year error component would be negative among those who persist, creating a potential downward bias in our estimate of the return to general experience.

We address endogenous responses to idiosyncratic shocks by including in all our specifications four indicator variables that are set to one if the observation is from a classroom that represents the

teacher’s last year teaching at the school in any classroom, in the current school-subject combination, in the current school-level combination, and in the current school-subject-level combination, respectively. These indicators capture the extent to which the year before an assignment change tends to exhibit particularly low performance, thereby preventing such dips from being fit by the experience profile parameters of interest. In addition to controlling for the most plausible dynamic response to health shocks, these dummies also control for the possibility that teachers who anticipate quitting put forth less effort in their final year (which could also bias downward the estimated general experience profile). Indeed, the coefficients on the dummies corresponding to the last year in the school-subject and school-subject-level (Appendix Table F.4) are negative and statistically significant.

However, other scenarios that produce violations of Assumption 1 might involve trends over time in error components rather than merely single-year idiosyncratic shocks, so that our “last year” indicators are inadequate controls. One particularly plausible mechanism stems from the possibility of heterogeneity in the gains to experience among teachers.²³ Since both our baseline and restricted specifications constrain the gains from general experience to be common to all teachers, any heterogeneity in rates of growth among teachers in the sample will be reflected in the teacher-year error component, ν_{rt} .

Thus, our context-specific experience profiles could be biased upward if teachers with faster than average growth rates are more likely to stay in the courses and levels they are teaching: the average value of the teacher-year error component ν_{rt} would be higher for higher values of subject-specific or level-specific experience. This might occur if rapidly improving teachers are rewarded with the opportunity to continue teaching their courses (while forcing others to adjust to changing classroom demand created by, say, teacher turnover or variation in student cohort size).

An analogous bias could be created by endogenous responses to school-year shocks. For example, teachers may be more likely to quit a declining school, thereby creating holes in subject or level offerings that other teachers must be forced to fill. In this case, the school-year error component ϕ_{st} would be positively correlated with levels of context-specific experience, leading to overestimates of the gains to experience.

²³ [Atteberry et al. \(2013\)](#) finds evidence of heterogeneous teacher growth in New York City.

We can test both of these hypotheses jointly by examining whether the trend in a teacher’s performance (relative to the estimated experience profile) predicts the teacher’s future teaching assignments. Indeed, such a test will also reveal the potential bias from *any* other sources of dynamic assignment patterns that involve a time trend in the composite error ϵ_{ct} within a teacher.

Specifically, we first identify all teacher-year combinations in which a teacher fails to teach any classroom in the following year. We then calculate and plot in Appendix Figure F.1a the average test score residuals across all classrooms of students taught by the teachers from these teacher-year combinations in the years leading up to their breaks from teaching (denoted t in event time). We see no evidence of any trend in teacher-year residuals in advance of the break from teaching. In order to distinguish quits/retirements from parental leave, Figure F.1e plots the same time path of teacher-year residuals leading up to the smaller sample of teacher-year combinations in which a teacher fails to teach in *any* future year in the sample. No obvious trend is observed.

We then perform the analogous exercise for changes in subject, level, and subject-level assignments. Specifically, for Figure F.1b (F.1f) we identify all teacher-subject-year combinations in which the teacher fails to teach any classrooms in the chosen subject in the following year (any future year), and plot the time path of average teacher-subject-year residuals leading up the change in subject assignment. Figures F.1c - F.1h plot the analogous trends in teacher-level-year and teacher-subject-level-year residuals leading up to breaks from teaching a given difficulty level or subject-level combination. None of the Figures F.1a-F.1h show any evidence of a significant trend in residuals preceding an assignment change that might suggest biases from dynamic reallocations of teacher assignments in response to unobserved shocks/input trends.²⁴

3.6.2 Testing for Dynamic Student Sorting

In this subsection, we focus our attention specifically on violations of Assumptions 1 and 2 that are caused by nonrandom student sorting. To gauge the possible severity of the problem, we implement a “backcasting” test in the spirit of Rothstein (2010) in which we replace class averages of students’

²⁴The point estimates that underlie these figures are presented in Appendix Table F.5.

contemporaneous test scores with class averages of their math standardized test scores from 7th grade.²⁵ The intuition behind the test is that if students were randomly assigned to teachers conditional on controls, current teacher identity or experience should not predict past student performance. To the extent that it does, part of the estimated gains to teacher experience could simply be capturing the ability of more experienced teachers to attract/be assigned to unobservably superior students.

The results of this exercise for the restricted specification 3.3 are presented in Appendix Table F.6. While the estimates are generally relatively small in magnitude, a number of estimates are statistically significantly different from zero, creating some cause for concern. A closer look, though, reveals that teachers with more general and subject-specific experience seem to be attracting students with inferior 7th grade math scores (conditional on 8th grade scores and the other controls), while teachers with more level-specific and subject-level specific experience seem to be attracting students with superior past test scores. Thus, the backcasting test suggests that the substantial gains to general and subject-specific experience reported above are, if anything, understated. By contrast, the gains to level-specific and subject-level specific experience could be slightly negative. Thus, these results do not undermine the qualitative conclusions of Section 3.5.2.

Furthermore, while such backcasting tests are well known in the literature and are valuable for flagging potential selection and sorting biases, recent research by Kinsler (2012) and Goldhaber and Chaplin (2015) suggest that these tests may find evidence of significant dynamic student sorting even where none exists. For example, suppose that classroom assignment in 9th or 10th grade is partially based on 7th grade test scores (perhaps because these test scores still affect principal or student beliefs about student ability), but that the part of persistent student inputs captured by 7th grade test scores is fully reflected in the included controls. In this case, current teacher assignments could significantly predict past test score noise or transitory student inputs,

²⁵These test scores are not included in our baseline specification because 7th test scores are missing for our first cohort (since they had already reached 8th grade the first year the statewide database was constructed). We wanted a consistent set of controls for all cohorts in our sample, and did not want to exclude our earliest student cohort, since their 1997 performance creates a baseline of productivity for the 1997 cohort of new teachers, which permits estimation of the gains to the 14th year of teacher experience and generally increases the precision of estimates of mid-career teaching (to which few cohorts of teachers contribute).

yet estimates of teacher value-added and gains from experience would nonetheless be unbiased.²⁶ Indeed, when we add 7th grade math and reading scores as controls as a robustness check in the next section, we find negligible changes in estimated gains from general and context-specific experience.

3.6.3 Evaluating Forecast Bias in Estimates of Context-Specific Teacher Talent

While the previous subsections have investigated several sources of potential bias in our estimated experience profiles, in this section we seek to determine the degree to which our estimates of teacher talent, the estimated fixed effects $\{\hat{\mu}_{srjl}\}$, properly capture the true talent contributions $\{\mu_{srjl}\}$. Following Chetty et al. (2014a), we do this by measuring forecast bias: the degree to which teachers' context-specific talent estimates from one partition of our data predict mean residual achievement in the same context in a second, left out partition. The implementation of our tests for forecast bias, which mirrors Chetty et al. (2014a), is described in detail in Section E.

We first test for forecast bias in our estimates of combined general and task-specific talent $\{\hat{\mu}_{srjl}\}$. This involves regressing differences in the performance of pairs of teachers within the same school-subject-level context from a left-out sample of classrooms on our posterior mean belief about the difference in the two teachers' talent in the chosen context. This empirical Bayes (EB) posterior belief is formed by multiplying the difference in estimated fixed effects $\hat{\mu}_{srjl} - \hat{\mu}_{sr'jl}$ from the primary sample by a reliability ratio that shrinks the estimated difference toward zero. If the estimated variance in teachers' talent contributions across randomly chosen test scores presented in Table 3.4 is valid, multiplying by this reliability ratio removes the attenuation bias created by sampling error in the fixed effect estimates $\{\hat{\mu}_{srjl}\}$ that would otherwise occur in a forecast regression of outcome differences from one sample on outcome differences in a second disjoint sample. Consequently, under the null hypothesis that the estimated (lower bound) teacher talent variance across tests is valid, the coefficient on the EB posterior mean from the forecast regression should converge in probability to 1.

²⁶Similarly, Chetty et al. (2016) point out that track-level, field-level, or school system-level shocks that are correlated across years could produce sampling error that is correlated across students' current and past classroom observations. This represents an additional mechanism by which backcasting tests could yield spurious "evidence" of bias.

The actual estimated regression coefficient (Appendix Table F.7, Column 1) is 0.825, with a standard error of 0.019. While this exercise reveals that our estimates of teacher talent can be used to forecast contributions to student performance out-of-sample fairly accurately, our estimator does not seem to be “forecast unbiased”. The most straightforward explanation for a coefficient below 1 is that our estimate $\hat{Var}(\mu_{stcl})$ slightly overstates the true variance in teacher talent contributions $\hat{Var}(\mu_{stcl})$, so that the reliability ratio we use in shrinkage overstates the degree of signal in the fixed effect differences. However, a couple of alternative explanations exist. First, the reliability ratio could also be overstated if we are underestimating the standard errors used to construct the estimated “noise”. Second, the subsample of school-teacher-subject-level combinations that satisfy the criteria for eligibility for the forecast sample (See Section E) might feature a slightly lower true variance in teacher talent contributions than the population.

While this test captures the model’s ability to consistently estimate the combined general and context-specific talent that a teacher contributes to a given context, the ability to improve the efficiency of teachers’ classroom assignments only depends on the model’s success in isolating and consistently estimating the context-specific components of teacher talent. Thus, we also construct two additional forecast tests that measure the degree to which our estimates of subject-specific and level-specific talent can forecast out-of-sample teachers’ subject-specific and level-specific comparative advantages, respectively.

Unlike our tests of the consistency of our combined talent estimates, which could be performed using differences among teachers who taught in the same school-subject-level context, evaluating our comparative advantage estimates requires measuring the degree to which difference-in-differences between teachers who taught the same two courses at the same school can be forecast. This necessitates restricting the forecasting sample to pairs of teachers who each taught multiple classes in the same two subjects within the same school-level combination (or, for the second test, both basic and honors in the same school-subject combination). Only 205 and 289 difference-in-differences exist on which to perform the forecast test for subject-specific and level-specific talent estimates, respectively. In essence, there is far less overidentifying variation available to test the model’s ability to detect true variation in subject-specific and level-specific talent.

The methodology for the context-specific forecast tests is otherwise perfectly analogous to the

forecast test for combined teacher talent. Difference-in-differences in residual mean test scores from among the left-out classrooms in the forecasted sample across teachers and either subjects or levels (conditioning on the same school-level or school-course environment as appropriate) are regressed on empirical Bayes estimates of difference-in-differences in the teachers' context-specific talent from the forecasting sample.

The regression coefficient on the forecasted difference-in-difference from the subject-specific forecast sample is 1.013, with a standard error of 0.242. Thus, while the point estimate suggests negligible forecast bias in estimates of subject-specific talent, the confidence interval is quite wide: only values below 0.539 can be ruled out with 95% confidence. Nonetheless, the test provides some reassurance that the kind of achievement data available to principals can provide some meaningful signal of subject-specific skill that might be used to guide classroom assignments.

The regression coefficient from the level-specific forecast regression is 0.456, with a standard error of 0.333. The point estimate indicates that our ability to infer level-specific talent is less strong than what our estimate of the true variance in level-specific talent would suggest. However, the test is severely underpowered: both 0 and 1 are within the 95% confidence interval. The large standard errors are partly due to the limited overidentifying variation just discussed, but are also attributable to the small estimated variance in level-specific skill: each classroom provides an extremely weak signal of level-specific skill relative to the “noise” stemming from the contributions of general teacher talent and other student and school inputs.

3.6.4 Further Robustness Checks

This subsection aims to provide a broader sense of the robustness of the main results to the array of difficult choices regarding specifications, variable definitions, and sample restrictions described in sections 3.2 and 3.4.

First, so far we have defined experience in a given context as the number of previous years in which the teacher taught at least one classroom in that context. This assumes that additional classes taught simultaneously in a context within a year (e.g. two periods of honors Biology classes) do not provide additional productivity value, which is based on the idea that teachers often have little time

to alter materials between classes in a given day. However, Appendix Table F.8 presents estimated experience profiles in which experience in each context is defined as the total number of classrooms taught in the chosen context in prior years. While the scales are difficult to compare, the results based on the classroom-based definition of experience are qualitatively very similar to those based on the year-based definition: substantial gains to general experience, moderate gains to the first few years of subject-specific experience, and negligible gains to level-specific and subject-level specific experience. Due to the near perfect correlation between year-based and classroom-based measures of experience, we are unable to determine which measure better captures the true accumulation of productivity gains from experience. Appendix Table F.9 shows that the decomposition of the variance in teacher talent is insensitive to the definition of experience.

Second, while we allow the permanent component of teacher productivity to be school-specific, to this point we have assumed that gains from general experience and from each dimension of context-specific experience retain their full value at new schools. Appendix Table F.10 presents estimated experience profiles based on the subsample of classrooms associated with teachers teaching in their first schools, where there is no concern about mismeasurement of experience stocks due to imperfect portability across schools. This subsample comprises 83.5 percent of our full sample of classrooms. The experience profiles remain essentially unchanged. Similarly, the decomposition of teacher talent for this subsample (Appendix Table F.11) is nearly identical to its full sample counterpart.

Third, Appendix Tables F.12 and F.13 present results from a specification in which we alter our controls for depreciation in experience-based human capital. Specifically, we replace indicators for whether the teacher taught the chosen subject, level, and subject-level (and whether the teacher taught at all) in the last year with linear controls for the number of years since having taught the relevant subject, level, or subject-level (or taught in any classroom). The estimates of the returns to experience are not sensitive to our handling of depreciation in experience-based human capital, and our depreciation controls are generally close to zero and statistically insignificant.

Fourth, Section 3.6.2 revealed that the identities of students' high school teachers can partially predict their prior 7th grade test scores, suggesting that 7th grade math and reading test scores might be valuable controls for student sorting. Thus, Appendix Table F.14 reports estimated experience profiles from a specification that includes class-averages of 7th grade math and reading

test scores as controls (and sets missing 7th grade test scores to the samplewide mean of zero). Inclusion of 7th grade math and reading scores has almost no impact on the estimated profiles. These results reinforce the idea that failure of a backcasting test need not imply substantive bias in estimates.

Fifth, the baseline and restricted specifications presented in Tables 3.5 and 3.6 impose that the returns to general and subject-specific experience are the same across fields. In Appendix Table F.15, we present separate estimates of general and subject-specific experience profiles for math, science, social studies, and English subjects. Comparing across columns, we see that general and subject-specific returns to experience are fairly similar across all four fields, providing support for the pooled specifications above. However, there is some variation in experience gains across fields. In particular, the gains to general experience appear highest in math, and the gains to subject-specific experience appear to be highest in science.

Sixth, up to this point we have combined years of experience 5 and 6, 7 through 10, and 11 and beyond into bins rather than introducing separate indicator variables for each year of experience. We did this because we expected gains from experience to slow down at higher levels of experience (as our estimates suggest they do), and combining multiple years into bins allows us to reap additional efficiency gains and identifying power (necessitated by the need to observe teachers' full teaching histories in order to construct their stocks of general and task-specific experience, which removes most well-experienced teachers from the sample). However, Wiswall (2013) points out that grouping experience levels into broad bins imposes arguably unrealistic restrictions on the experience profile that can potentially produce substantial bias. To address this concern, Appendix Table F.16 presents estimated general and context-specific experience profiles from a version of the restricted specification in which indicators are included for years 1 through 14 of experience.²⁷ While the estimates become prohibitively noisy beyond six or so years of experience, the results for the first several years of experience are extremely similar to those presented in Table 3.6 across all four dimensions of experience, suggesting that pooling multiple years of experience

²⁷When the full set of dummies is introduced into the baseline specification, the results become nonsensical, with enormous offsetting positive and negative effects across dimensions. This is not surprising, as combining multiple years into bins was helping to break the collinearity between the various dimensions of experience, so that the overfitting/collinearity problem discussed in Section 3.5.2 now becomes even more severe.

into experience category indicators is not generating substantial bias, at least for the unpooled experience categories.

Along the same lines, Appendix Table F.18 displays predicted values for the first ten years of experience in each dimension from a specification in which the set of indicators for each number of years of experience is replaced by a quartic in each experience dimension (general, subject-specific, level-specific, and subject-level specific). The gains to each dimension of experience are again similar, illustrating that a smoother, more parsimonious specification can still capture the basic qualitative results.

Finally, both the baseline and restricted specifications impose that the separate components of experience are additively separable in the education production function:

$$d(exp^{gen}, exp^j, exp^l, exp^{jl}) = d^{gen}(exp^{gen}) + d^j(exp^j) + d^l(exp^l) + d^{jl}(exp^{jl}) \quad (3.12)$$

However, general experience and different dimensions of context-specific experience may interact with one another. For example, perhaps students only learn if the teacher has developed effective ways to both explain a subject's content *and* maintain control of the classroom. Lectures that deliver content effectively may require subject-specific experience, whereas classroom control skills may be learned through general or level-specific experience. Alternatively, perhaps a teacher can keep student attention by either having exceptional command of the content *or* by having excellent classroom control skills, in which case the different components of experience would be substitutable rather than complementary.

We relax the additive separability assumption by estimating a “full” specification that captures the contribution of experience to teacher productivity via a non-parametric function of the four experience components:

$$Y_{ct} = X_{ct}\beta_{jl} + \delta_{sjl} + \mu_{srjl} + d(exp_{rt}^{gen}, exp_{rt}^j, exp_{rt}^l, exp_{rt}^{jl}) + \epsilon_{ct} \quad (3.13)$$

We implement this specification by replacing the four dimension-specific experience profiles with a full set of four-dimensional experience cell fixed effects.²⁸ This specification is isomorphic in

²⁸For example, the vector of experience stocks $(exp^{gen}, exp^j, exp^l, exp^{jl}) = (2, 1, 1, 1)$ is captured by a different

structure to a model with worker and firm fixed effects. Since the estimated experience cell fixed effects are measured with considerable sampling error, to better reveal the underlying structure of the experience contributions we smooth estimates for each experience cell by using a normal kernel to give weight to “nearby” estimates.²⁹

We then take partial derivatives of this smoothed non-parametric experience production function with respect to each dimension of experience and integrate over these dimension-specific partial derivative functions to construct a set of standard experience profiles analogous to those from our additively separable baseline and restricted specifications. Section D.2 describes this procedure in further detail.

The results of this exercise are displayed in Table F.19, while Table F.20 displays the corresponding marginal effect estimates from a “restricted” version that replaces school-teacher-subject-level fixed effects with school-teacher fixed effects. Compared to the additively separable results from Table 3.6, the results in Table F.20 feature quite similar general and subject-level experience profiles, but somewhat larger gains to both subject-specific and level-specific experience. Overall, though, accounting for possible misspecification from ignoring interactions among experience components does not change the basic qualitative conclusion that the bulk of the gains from experience stem from general and subject-specific experience.³⁰

indicator variable than (2, 1, 2, 1).

²⁹D.1 provides a more detailed explanation of this smoothing procedure.

³⁰Previous versions of this paper attempted to use the full specification to characterize the nature of complementarity present in the experience production function. While such attempts produced suggestive evidence that general, subject, and level experience are substitutes rather than complements, identification and estimation of the degree of complementarity places extremely strong demands on the data, so that the results were both noisy and fragile. Thus, we have chosen to remove the full discussion of sub/supermodularity of the production function in this version.

3.7 Projecting the Achievement Gains from Efficient Use of Context-Specific Teacher Experience and Talent

3.7.1 Methodology

The moderate variance in subject- and level-specific time-invariant productivity differences among teachers, combined with the estimated gains to subject-specific experience, suggest that fully exploiting a teaching staff's task-specific human capital could potentially generate non-trivial efficiency gains. In this section we develop a set of counterfactual simulations to gauge the magnitude of the performance gains that could be achieved statewide if each principal exploited the full value of the stocks of task-specific experience and talent of the members of his or her teaching staff.

To see how such simulations might be implemented, consider the allocation of teachers to classrooms that takes place at a particular school in a particular field over the set of years in our sample. Ideally, we would solve the dynamic problem of choosing sequences of yearly allocations to maximize the average test score performance over the entire sample (and perhaps beyond). However, the state space of such a dynamic problem is prohibitively large: it must include, for each teacher in the school, both the teacher's stock of task-specific experience as well as posterior beliefs (with corresponding precisions) about the teacher's talent in each subject and level.

Consequently, we instead simulate the dynamic effects of re-solving each year the static optimization problem in which the expected average test score for the year is maximized, taking the set of classrooms and teachers to be matched in the chosen year as exogenously given at the start of the year. Four-dimensional experience stocks are then updated for the next year based on the efficient static allocation. While this approach necessarily understates the true gains to dynamic optimization, it represents an allocation rule that principals can automatically implement each fall with minimal computational burden and without making any projections about enrollment and teacher attrition. By evaluating the dynamic implications of static optimization, we can ensure that the short-run efficiency gains from implementing the statically optimal allocation are not undermined by long-run efficiency losses.

Even static optimization, however, requires specifying the principal's belief about each teacher's

time-invariant task-specific productivity for each subject-level combination to which the teacher could potentially be assigned. Thus, we calculate empirical Bayes posterior beliefs about each teacher’s task-specific talent based on our school-teacher-subject-level fixed effect estimates and their standard errors, and use these for any school-teacher-subject-level combinations that are observed in our sample. We assign task-specific productivities of 0 (the population mean) to any school-teacher-subject-level combination that we do not observe.

These posterior beliefs are designed to make efficient use of the information about teacher comparative advantages contained in the student test score data, given our assumed achievement production function. However, a major concern is that the principal may have information about teachers’ subject-specific or level-specific talent that is not reflected in the test score data, and is therefore unobserved by the econometrician. Such information might be derived from classroom observations or from knowledge of the teacher’s college preparation (e.g. a biology major might be likely to have a comparative advantage in biology relative to chemistry). If such additional sources of principal information exist, then allocations that are optimal based on the posterior beliefs we calculate may identify spurious efficiency “gains” in which teacher assignments that were driven by the unobserved component of principal’s information are altered to better fit the noise in our empirical Bayes estimates of task-specific talent, and thus would in fact represent achievement *losses*.

In light of this concern, we compare our simulated “optimal” allocations to two different baselines representing different informational assumptions. The first baseline consists of the achievement contribution of actual teacher assignments in our sample under the assumption that the information available to the principal at the time of allocation is a subset of the information contained in our entire sample of test scores for all public school teachers in North Carolina. Under this assumption, the gains we identify from teacher reallocations should be correct in expectation; we are as likely to understate as to overstate the gains from our alternative allocation.³¹

³¹ Given that the fixed effect estimates are based on the entire sample, one could argue that they are partially based on information (teachers’ average test score performance from future years) that principals cannot possibly have observed at the time of allocation, so that the gains we compute overstate gains from a feasible allocation algorithm even if principals do not have other sources of information on teachers’ task-specific talent. However, for highly experienced teachers that are included in only a subset of our simulations (but would make up a substantial fraction of the principals actual staff), the principal will generally have many years of test-score data on which to

However, since this assumption may cause us to overstate (possibly dramatically) the potential efficiency gains from effective use of test-score-based information about task-specific experience and talent, we also compare the achievement gains from our “optimal” allocations to a baseline in which teachers are randomly allocated to classrooms within field. This random baseline allows the reader to gauge the potential importance of utilizing information about task-specific experience and talent contained in test score data without making any assumption about the degree to which this and other sources of information are already being used by principals. Principals and other administrators may simply wish to know whether it is worth the time and effort to track task-specific experience and generate beliefs about task-specific talent and to incorporate these pieces of information into classroom assignments, or if instead they should allocate classrooms based on other objectives that may be nearly orthogonal to the short run maximization of test scores (e.g. minimizing parent dissatisfaction).

To ensure that the simulation captures feasible reallocations, we hold fixed the number of classrooms of each subject-level combination at the levels that actually prevailed at each school in each year. Furthermore, we also hold fixed the total number of classrooms taught by each teacher in each year, since principals may have been constrained in the workload they could assign to their more experienced teachers.³² This counterfactual simulation can be rewritten as a binary integer programming problem. The formal presentation of the problem is located in [F](#).

Since our estimated gains from general and task-specific experience are based on only the 11 tested subjects, our simulations only consider efficiency gains from reallocating classrooms in which the tested subjects were taught. We also do not reallocate classrooms in which English 1 was taught, since this is the only tested subject in English. In addition, because we do not observe the full

base posterior beliefs, so that their posterior beliefs may closely correspond to our empirical Bayes posterior beliefs. Another possible approach would have been to calculate posterior beliefs for each teacher in each year based on their performance record up to that date. This would be quite computationally costly for us, since it requires re-estimating the model (and calculating standard errors) for each year in our sample, but would likely be feasible for an actual school that is allocating only a handful of teachers. Thus, even our richer static optimization program could be fairly easily implemented by any school, given accurate records on the teachers’ past course assignments and student performance.

³²For example, these teachers may also have been teaching untested classes, or performing other valuable services to the school, such as lunchroom monitoring, advising student clubs, or coaching student athletic teams.

teaching histories of any teacher who began teaching before the sample begins in 1995, for some of our simulations, we do not reallocate the classrooms taught by such teachers; for other simulations, we impute the full teaching histories for such teachers. We use the estimates from the full specification in equation (3.13) for both the posterior beliefs about context specific talent and the predicted contributions of each four-dimensional vector of experience stocks.³³

Our simulation procedure captures the gains that could have been reaped by the end of each year had the principal maximized the value of context-specific experience and context-specific talent in each school starting in 1995 (the first year of the sample). However, estimates in the first few years of the sample conflate the fact that past rotations have limited potential gains from re-optimizing with the fact that relatively few teachers are being reallocated.³⁴ Thus, we focus on efficiency gains among classrooms assigned in the last 5 years of the sample, when a substantial fraction of teachers are eligible for reassignment. We do not extend our simulations beyond the last year of the sample, so that the gains we report may not fully capture the very long run (steady state) gains from repeated optimization. We hold the allocation of teachers to schools fixed (thus ignoring any possible effects of classroom reassignments on teacher turnover), and we continue to assume that context-specific experience is fully portable across schools.

We also compare the results of the “dynamic” simulation to a fully static simulation that solves the binary integer programming problem in each year t holding fixed observed teacher assignments up through $t - 1$. These results reflect the payoff to the first year of optimal static reallocation. The static simulation serves to illustrate the decomposition of gains into the part stemming from initial reassignment to better match teachers’ context-specific experience and talent to the courses

³³We smooth the nonparametrically-estimated experience function to a greater degree for the simulations to ensure that our simulated efficiency gains do not stem from better exploiting the sampling error portion of the estimated returns to experience. We use a bandwidth (variance on a normal PDF) of 5 to smooth the estimates used for the simulation. In theory, the appropriate smoothing represents a delicate balance: smoothing too little creates the possibility of spurious gains from better fitting sampling error in estimates, while smoothing too much also removes the signal. Indeed, complete smoothing would make the productivity of each experience cell identical, and would therefore eliminate the possibility of any gains from better use of teacher experience stocks. In practice, however, we have found that bandwidth choices between 2 and 10 yield very similar estimates.

³⁴This is because we do not observe the classroom assignment histories for the vast majority of the teachers in the first few years.

they teach and the part stemming from longer run gains associated with the specialization of the teacher work force.

3.7.2 Results from Counterfactual Simulations

The bar charts in Figure 3.4 present the student-weighted average expected test score gain from optimal reallocation among all school-year combinations for both the single-year “static” simulation (Panels C and D) and the “dynamic” simulation in which static re-assignments affect the following year’s experience stocks (Panels A and B). The numerical values that correspond to the bars in Figure 3.4 are presented in Table 3.8.

The results in Figure 3.4 are reported separately by whether the baseline is the actual allocation observed in the sample (Panels A and C) or an allocation in which teachers are randomly assigned to classrooms within field (Panels B and D). In addition, because the scope for efficiency gains from matching and specialization increases in the size of the teaching force, achievement gains are also presented separately by number of teachers in the school-field-year combination eligible to be reallocated (i.e. the number who taught at least one classroom in that school-field-year combination in the actual data for whom the full teaching history is observed).³⁵

While optimal reallocations were implemented separately by field, the results displayed in Figure 3.4 pool the classroom-level gains across the three fields (math, science, social studies). We pool the results because there was surprisingly little heterogeneity in simulated gains from reallocation across fields (See Appendix Table F.21 for the disaggregated results). In each panel of Figure 3.4, the height of the rightmost bar in each set of three bars represents the total simulated per-student standardized test score gain from optimally allocating teachers to classrooms, while the heights of the leftmost and middle bars decompose this total per-student gain into the components stemming from better (or worse) use of teachers’ context-specific experience and context-specific talent, respectively.

³⁵In the case where only one teacher is observed teaching all of the courses in the field, there can be no gains from teacher reallocation. Thus, school-field-years featuring only one teacher are omitted from the simulations presented in Figure 3.4.

We focus first on Panel A, which displays results from our “dynamic” simulations in which the actual allocation observed in North Carolina is used as a baseline. These results indicate that better use of context-specific talent in particular has the potential to reap non-trivial efficiency gains. Specifically, the total gains relative to the actual allocation from better use of context-specific teacher productivity grow from .017 test score standard deviations for school-year-fields in which only two teachers are eligible to be reallocated to .033 for four-teacher fields and .044 standard deviations for school-year-fields featuring eleven or more eligible teachers.³⁶ Moreover, these total gains derive almost entirely from more efficient use of teachers’ task-specific talent, while gains from better use of teacher-specific experience are negligible and in some cases slightly negative.

If instead the random allocation is used as a baseline (Panel B), two-teacher fields reap efficiency gains of .025 standard deviations, while four-teacher fields produce gains of .044 standard deviations and fields with eleven or more teachers produce gains .054 standard deviations. Generally speaking, about 20% of the gains relative to the random allocation comes from effective use of context-specific experience rather than talent; .05, .010 and .015 of the total per-student test score gain can be attributed to better exploiting teacher subject-specific and level-specific experience for school-year-fields with two teachers, four teachers, and eleven or more teachers eligible for reallocation, respectively. The combined results in Panels A and B suggest that while effective use of teachers’ stocks of context-specific experience could be an important source of efficiency gains in some contexts, North Carolina principals already seem to be effectively exploiting the context-specific experience of their staffs, possibly even at the expense of subject-specific and level-specific teacher talent.

Panels C and D display the corresponding results for the “static” simulations, in which teacher assignments up until time $t - 1$ are held fixed when choosing simulated classroom allocations at t . They reveal that nearly all of the long-run gains from optimal reallocation of teachers are reaped in the first period of reallocation. This is not surprising given the small fraction of the total efficiency

³⁶This pattern is mirrored in the fraction of classrooms whose assigned teacher in the simulation differs from the one observed in the data. 30.3, 44.9, and 51.9 percent of classrooms in the math field with two, four, and 11+ teachers have their original teachers reassigned in the dynamic simulation. The corresponding percentages are 25.3, 40.4, and 45.6 for the static simulation. Appendix Table F.22 presents the full set of reallocation rates from our simulations.

gains in the dynamic simulations attributable to better use of task-specific experience.

Note that if principals have very precise information about task-specific talent at the time of hire, then there is no tension between maximizing the contributions of task-specific experience versus task-specific talent: teachers can be assigned when hired to the courses in which they have the strongest comparative advantages, and then can continue to teach these courses, building up the relevant task-specific experience. However, for principals that have minimal information about teachers' context-specific talent at the time of hire, our estimates suggest that the degree to which teachers should be rotated among courses is likely to depend strongly on a school's teacher turnover rate. For schools with very low turnover rates, the variance in context-specific talent is sufficiently large that principals might find it worthwhile to rotate teachers for several years in order to learn the set of course assignments that best utilize task-specific talent. However, for schools with high turnover rates, the signal about task-specific talent received from a small number of classrooms is sufficiently coarse that the knowledge necessary to benefit from superior allocation of task-specific talent cannot be gathered in time for it to be valuable; by contrast, the productivity gains from the first two years of subject-specific experience are reasonably large, and can be reaped even among teachers who are only likely to stay for three or four years. This logic suggests that high turnover schools are likely to be better off minimizing the degree to which teachers are rotated among courses.

In the absence of an analytical solution to the full dynamic problem, a more precise characterization of the optimal amount of teacher rotation requires simulating test score contributions from alternative rotation strategies for a variety of parameter combinations governing, for example, turnover rates, principal information, teaching loads, and the number of distinct subjects, levels, and courses. We leave such an extensive simulation exercise for future work. However, we wish to emphasize that each individual school likely faces fixed and known values of many of these remaining parameters, so that the estimated task-specific experience profiles and underlying variances in subject-specific and level-specific talent presented in this paper provide the information necessary for school administrators to perform their own customized simulations to guide their classroom assignment decisions.

In Appendix Table F.23, we also display results from simulations in which all observed teachers who

taught the tested courses are eligible for reallocation. We impute context-specific experience stocks for those teachers whose full teaching history is not observed based on the distribution of context-specific experience among the most experienced teachers whose full histories are observed. Adding in the full roster of teachers reduces dynamic gains relative to the actual allocation to .005 standard deviations for two-teacher fields, .014 standard deviations for four-teacher fields, and .025 for fields with eleven or more teachers (though note that 29% of the school-year fields in the full sample feature 11+ teachers, relative to 2% for the complete history subsample). These smaller simulated gains indicate that principals might make better use of their experienced teachers' context-specific talent, suggesting that they may learn teachers' comparative advantages slowly. When the random baseline is used instead, the corresponding gains are .011 for two-teacher fields, .027 standard deviations for four-teacher fields and .042 for fields with eleven or more teachers.

On one hand, these magnitudes are clearly not large enough to dramatically shift the distribution of student achievement; a .025 standard deviation test score gain is only enough to move an average student from the 50th to the 51st percentile of the state test score distribution. However, a number of other considerations suggest a more optimistic interpretation of these efficiency gains.

First, note that these gains are virtually costless: no change in existing staff is required, and all teaching loads are held fixed. It is rare to find the potential for across-the-board gains from policy changes that require so little upheaval.

Second, given that the vast majority of the test-score variation is within classes, most other school-level policies are likely to have a similarly-sized impact. For example, consider a policy that aims to identify and replace the worst 10 percent of teachers with new hires. Using the estimates from Table 3.4, the expected contribution of a randomly chosen teacher below the 10th percentile of general skill is -.22 test score standard deviations, so that if such teachers teach only 10 percent of students, average test scores would increase by 0.022 standard deviations even under the optimistic assumption that replacement teachers were of average quality.

Third, note that the vast majority of students are taught in high schools that feature seven or more teachers in a field. Furthermore, classrooms were only reallocated in tested courses, so that, for example, teachers who only taught calculus were not available for reallocation. Thus, the largest

efficiency gains from our simulations are probably the relevant gains in most situations, and in fact may still be underestimates for most large schools.

Finally, these average gains conceal considerable heterogeneity in potential gains among schools. Consider the specification that incorporates task-specific talent, reallocates only teachers with fully observed teaching histories, and uses the observed allocation as the baseline. Focusing on schools with fields that generally feature seven or more teachers and averaging across fields and years, the mean dynamic gain from optimal reallocation among the 10 percent of schools featuring the smallest gains is only .004 standard deviations, while schools among the top decile of the distribution of dynamic gains are predicted to enjoy test score increases of .047 standard deviations on average. Thus, there seem to be a non-trivial subset of schools that might be able to reap substantial gains simply from changing their teacher assignment mechanism.

On the other hand, several additional caveats and limitations of our simulations should be noted. First, recall that the projected gains relative to the actual allocation rely on the questionable assumption that the principal does not have alternative sources of information beyond what is reflected in the full sample of test scores. Second, we are unable to evaluate the extent to which any achievement gains from an alternative teacher assignment mechanism would also contribute to or detract from other important non-test score student or school outcomes. Furthermore, because we do not allow our simulated classroom assignments to affect teacher turnover, the simulated efficiency gains could overstate even the true achievement gains if, for example, good teachers have a taste for variety, and quit more frequently if they are forced to teach the same subject-level combination repeatedly.³⁷ Similarly, our data do not permit us to estimate gains (or losses) to high levels of general and context-specific experience. It may be that some of the excess rotation of teachers away from their comparative advantages is necessary to prevent burnout or human capital depreciation among the most senior teachers. This might also lead us to overstate potential gains from optimal reallocation.³⁸

³⁷However, [Ost and Schiman \(2015\)](#) suggests that the opposite is true in the elementary school context: teachers who rotate more frequently among grades exit schools at a higher rate.

³⁸Note, though, that this scenario could also cause us to understate the gains to reallocation, since we have essentially assumed away any experience-based gains from reallocation among very experienced teachers by assigning the same experience productivity values to all levels of experience beyond 10 years.

3.8 Conclusions

This paper introduces and implements a method for decomposing worker productivity into task-specific and general components of both experience and persistent talent. For high school teachers, about a third of the productivity gains from experience are specific to the subjects to which a teacher has been assigned, while about 74% of the variance in experience-invariant talent is portable across all courses. Nonetheless, our simulations provide suggestive evidence that existing allocations of teachers to classrooms in public high schools might be failing to exploit the variation in subject-specific and level-specific human capital that does exist, suggesting the potential for efficiency gains of around .02-.03 student test score standard deviations on average, with larger gains for some schools.

Note, however, that the results of the decomposition we estimate may not generalize to other occupations or even to alternative definitions of teachers' tasks. In particular, the tasks we consider are still fairly similar in scope. For example, we might observe greater variation in task-specific talent among teachers if we included serving as a high school athletic coach as one of a teacher's tasks. Similarly, developing students' cognitive and non-cognitive skills might represent two different tasks facing a teacher even within a given classroom context.³⁹

The methodology, however, does generalize: a similar decomposition may be estimated in any context in which worker productivity may be measured at the task level and where the blend of tasks changes over time. Indeed, there are many other organizational contexts in which we might also expect productivity to reflect a mix of general and task-specific talent as well as general and task-specific experience, and in which the nature of this production function may not be easily observable by employers or managers. A company employing a sales team to sell different products to different types of clientele, for example, might have both the wherewithal and the need to implement our decomposition.

³⁹For instance, teachers who are effective at teaching abstract concepts may not be effective at handling student emotional crises.

3.9 Tables and Figures

Table 3.1: Effect of Sample Restrictions on Sample Composition

	Full Sample	Regression Sample
	(1)	(2)
School-Year Averages		
Enrollment	1,346.0 (654.5)	1,362.3 (646.4)
# Teachers	23.1 (8.9)	23.3 (8.7)
Teacher-Year Averages		
General Experience	4.96 (3.65)	3.28 (3.01)
Subject Experience	3.78 (3.30)	2.37 (2.48)
Level Experience	4.70 (3.55)	3.11 (2.91)
Subject-Level Experience	3.42 (3.11)	2.17 (2.35)
Classes Taught Per Year	3.44 (1.52)	3.34 (1.50)
Unique Subj./Lvl. Taught Per Year	1.67 (0.70)	1.63 (0.67)
Student-Year Averages		
Standardized Subject Test	0.041 (0.662)	-0.024 (0.636)
Fraction of White Students	0.667 (0.270)	0.641 (0.276)
Fraction of Black Students	0.259 (0.252)	0.278 (0.258)
Fraction of Other Students	0.074 (0.100)	0.081 (0.104)
8th Grade Standardized Reading Scores	0.095 (0.975)	-0.024 (0.614)
8th Grade Standardized Math Scores	0.075 (0.976)	-0.030 (0.651)
N (Aggregated Classroom Observations)	207,951	61,993

Notes: Student-test-weighted means and standard deviations (in parentheses) of classroom observations are reported for each sample. *Full Sample* includes all classroom observations with valid values for the variables in this table (i.e. current and 8th grade test scores, subject and level designation, race variables, teacher experience, class size, and grade). *Regression Sample* includes only classroom observations that satisfy the more extensive set of sample restrictions described in Section 3.4. The most important restriction is that the full history of course assignments must be observed for the teacher of the classroom. The *School-Year Averages* for the Regression Sample in Column (2) present the school-average student enrollment and teaching staff size from the *Full Sample*, but for the classrooms represented in the *Regression Sample* (now weighted by the number of student-tests in the regression sample). If we only count the subset of students and teachers that actually contribute an observation to our Regression Sample, student-test-weighted school means of enrollment and number of teachers are 504.9 and 8.4, respectively.

Table 3.2: Teacher Mobility Across Subjects: Regression Sample

		Math			Science				Social Studies			English
		Algebra 1	Algebra 2	Geometry	Biology	Chemistry	Physics	Physical Sciences	Civics	E/L/P	U.S. History	English
Math	Algebra 1	1,860	749	742	26	18	16	37	11	12	13	33
		1.000	0.403	0.399	0.014	0.010	0.009	0.020	0.006	0.006	0.007	0.018
	Algebra 2	749	1,078	533	4	9	14	14	1	3	3	0
		0.695	1.000	0.494	0.004	0.008	0.013	0.013	0.001	0.003	0.003	0.000
	Geometry	742	533	1,142	8	3	10	6	1	3	3	4
Science		0.650	0.467	1.000	0.007	0.003	0.009	0.005	0.001	0.003	0.003	0.004
	Biology	26	4	8	1,472	185	69	525	7	24	20	26
		0.018	0.003	0.005	1.000	0.126	0.047	0.357	0.005	0.016	0.014	0.018
	Chemistry	18	9	3	185	554	112	307	0	0	1	1
		0.032	0.016	0.005	0.334	1.000	0.202	0.554	0.000	0.000	0.002	0.002
Soc. Stu.	Physics	16	14	10	69	112	243	165	0	0	2	0
		0.066	0.058	0.041	0.284	0.461	1.000	0.679	0.000	0.000	0.008	0.000
	Physcial Sciences	37	14	6	525	307	165	1,151	6	24	15	21
		0.032	0.012	0.005	0.456	0.267	0.143	1.000	0.005	0.021	0.013	0.018
	Civics	11	1	1	7	0	0	6	904	279	412	12
English		0.012	0.001	0.001	0.008	0.000	0.000	0.007	1.000	0.309	0.456	0.013
	E/L/P	12	3	3	24	0	0	24	279	952	414	52
		0.013	0.003	0.003	0.025	0.000	0.000	0.025	0.293	1.000	0.435	0.055
	U.S. History	13	3	3	20	1	2	15	412	414	1,235	36
		0.011	0.002	0.002	0.016	0.001	0.002	0.012	0.334	0.335	1.000	0.029
English	English	33	0	4	26	1	0	21	12	52	36	2,162
		0.015	0.000	0.002	0.012	0.000	0.000	0.010	0.006	0.024	0.017	1.000

Notes: E/L/P denotes Econ/Law/Politics. The top entry in the (i,j)-th cell is the number of teachers who are observed teaching in both the i-th and the j-th subject (not necessarily in the same year). The bottom entry of the (i,j)-th cell is the fraction of teachers ever observed teaching the i-th subject who are also observed teaching the j-th subject at some point during the sample.

Table 3.3: Teacher Mobility Across Math Subject-Level Combinations: Regression Sample

		Algebra 1		Algebra 2		Geometry	
		Low	High	Low	High	Low	High
Algebra 1	Low Level	1,855	27	676	331	678	315
		1.000	0.015	0.364	0.178	0.365	0.170
	High Level	27	32	18	10	14	7
		0.844	1.000	0.563	0.313	0.438	0.219
Algebra 2	Low Level	676	18	966	341	451	194
		0.700	0.019	1.000	0.353	0.467	0.201
	High Level	331	10	341	453	202	118
		0.731	0.022	0.753	1.000	0.446	0.260
Geometry	Low Level	678	14	451	202	1,053	368
		0.644	0.013	0.428	0.192	1.000	0.349
	High Level	315	7	194	118	368	457
		0.689	0.015	0.425	0.258	0.805	1.000

Notes: The top entry in the (i,j)-th cell is the number of teachers who are observed teaching in both the i-th and the j-th subject- level (not necessarily in the same year). The bottom entry of the (i,j)-th cell is the fraction of teachers ever observed teaching the i-th subject-level who are also observed teaching the j-th subject-level at some point during the sample.

Table 3.4: True Variances in Fixed Effects (Using Year-Based Measure of Teacher Experience with the Baseline Specification)

	Lower Bound		Intermediate		Upper Bound	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0236	0.154	0.0467	0.216	0.0605	0.246
General Talent	0.0175	0.132	0.0368	0.192	0.0506	0.225
Subj-Lvl Combos	0.0061	0.078	0.0099	0.099	0.0099	0.099
Sch-Lvl-Tch Combos	0.0197	0.140	0.0407	0.202	0.0545	0.234
Subject Talent	0.0039	0.063	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0215	0.147	0.0433	0.208	0.0571	0.239
Level Talent	0.0021	0.045	0.0034	0.058	0.0034	0.058
Subject-Level Talent	0.0001	0.011	0.0005	0.023	0.0005	0.023

Notes: Standard errors are clustered at the teacher level. *Lower Bound* estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.3.2 for details.

Table 3.5: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Baseline Specification)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.066*** [0.017]	0.014 [0.015]	-0.006 [0.015]	0.000 [0.013]	0.074*** [0.006]
2 yrs	0.085*** [0.025]	0.033* [0.023]	-0.006 [0.022]	-0.013 [0.020]	0.099*** [0.010]
3 yrs	0.090*** [0.030]	0.049** [0.029]	-0.001 [0.028]	-0.027 [0.026]	0.110*** [0.014]
4 yrs	0.097*** [0.035]	0.053* [0.033]	-0.004 [0.033]	-0.033 [0.031]	0.113*** [0.018]
5-6 yrs	0.097*** [0.040]	0.056* [0.039]	0.008 [0.037]	-0.044 [0.036]	0.116*** [0.023]
7-10 yrs	0.113*** [0.046]	0.046 [0.046]	-0.006 [0.044]	-0.054 [0.043]	0.098*** [0.031]
11-14 yrs	0.093** [0.054]	0.067 [0.059]	0.024 [0.055]	-0.099** [0.060]	0.085** [0.046]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. The regression includes school-teacher-subject-level fixed effects, calendar year fixed effects, and a vector of classroom observable characteristics with subject-level-specific coefficients. The regression also includes controls for teacher workload (number of current class periods and number of distinct subject-levels taught) and depreciation of experience capital (indicators for whether the teacher taught a class in the current subject, level, subject-level, or taught at all last year, as well as analogous indicators for teaching in each context two years ago). Finally, the regression also includes controls for decreasing effort/productivity shocks in the year prior to an assignment change (indicators for whether the current year is the final time the teacher taught the subject, level, subject-level associated with the observation, as well as whether the current year is the teacher's final year of teaching high school in North Carolina. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table 3.6: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.065*** [0.011]	0.014* [0.009]	-0.003 [0.010]	0.013* [0.008]	0.089*** [0.004]
2 yrs	0.085*** [0.014]	0.023** [0.012]	-0.004 [0.012]	0.014* [0.010]	0.118*** [0.006]
3 yrs	0.093*** [0.016]	0.036*** [0.014]	-0.007 [0.014]	0.008 [0.012]	0.131*** [0.007]
4 yrs	0.101*** [0.018]	0.041*** [0.015]	-0.011 [0.016]	0.007 [0.014]	0.138*** [0.008]
5-6 yrs	0.103*** [0.019]	0.041*** [0.017]	-0.002 [0.017]	0.009 [0.015]	0.152*** [0.009]
7-10 yrs	0.114*** [0.022]	0.025 [0.021]	-0.008 [0.020]	0.006 [0.019]	0.138*** [0.012]
11-14 yrs	0.107*** [0.028]	0.027 [0.038]	0.027 [0.028]	-0.019 [0.041]	0.141*** [0.026]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to the specification in equation (3.3) in which the school-teacher-subject-level fixed effects $\hat{\mu}_{srjl}$ from Equation (3.2) are restricted to be common across subject-levels (i.e. replaced by school-teacher effects). Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table 3.7: Effect of Years of General and Subject-Specific Experience on Student Test Scores (Restricted Specification with Level and Subject-Level Experience Additionally Constrained to 0)

	Restricted Specification w/ Lvl. & Subj.-Lvl. Exp. Gains Constrained to 0			Standard Specification
	General	Subject	Combined	“General”
	(1)	(2)	(3)	(4)
1 yr	0.063*** [0.007]	0.025*** [0.006]	0.088*** [0.004]	0.084*** [0.004]
2 yrs	0.081*** [0.009]	0.036*** [0.008]	0.118*** [0.005]	0.113*** [0.005]
3 yrs	0.087*** [0.010]	0.045*** [0.009]	0.133*** [0.007]	0.127*** [0.006]
4 yrs	0.092*** [0.011]	0.048*** [0.010]	0.140*** [0.007]	0.136*** [0.007]
5-6 yrs	0.100*** [0.012]	0.050*** [0.011]	0.151*** [0.009]	0.148*** [0.008]
7-10 yrs	0.107*** [0.014]	0.032** [0.014]	0.139*** [0.011]	0.148*** [0.010]
11-14 yrs	0.124*** [0.020]	0.020 [0.025]	0.143*** [0.023]	0.157*** [0.016]

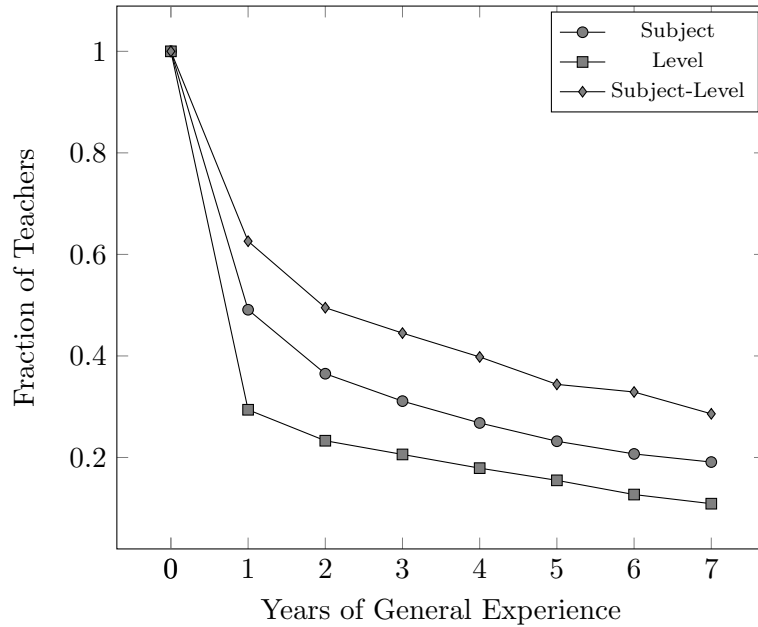
Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to the specification in equation (3.3). Columns (1-3) report results from imposing on the Restricted Specification the additional restrictions that gains from level-specific and subject-level-specific experience are constrained to be 0: $d^l(exp) = 0$ and $d^{jl}(exp) = 0 \forall exp$. Column 4 reports results from imposing the further restriction that $d^j(exp) = 0 \forall exp$, for ease of comparison with with standard experience profiles estimated in the literature. Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom’s teacher taught at least one class at all (Col. 1 & 4) or in the subject (Col. 2) associated with the current classroom observation. Column 3, entitled *Combined*, captures the combined predicted contribution of both dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table 3.8: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

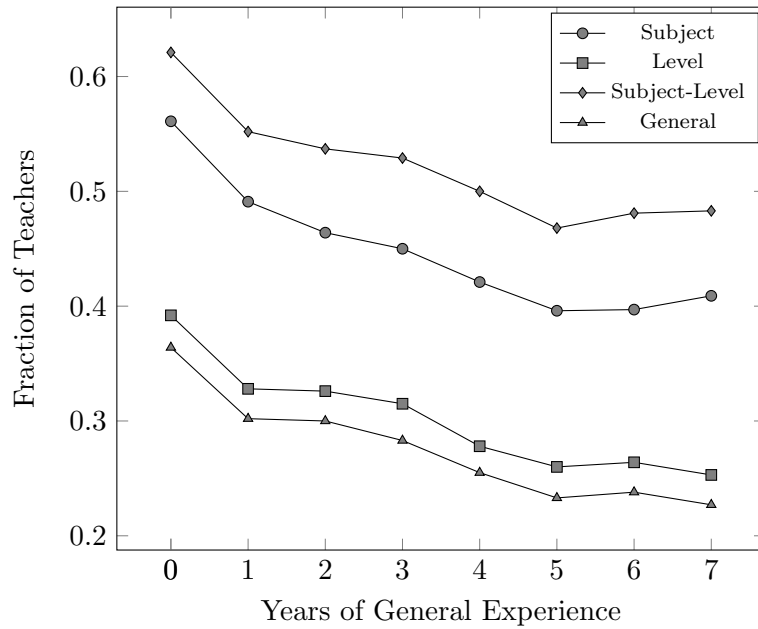
	Eligible Teach.	Static		Dynamic	
		Actual (1)	Random (2)	Actual (3)	Random (4)
2	Total	.017	.025	.017	.025
	Talent	.018	.021	.018	.020
	Exper.	-.001	.004	-.001	.005
3	Total	.026	.035	.027	.039
	Talent	.027	.030	.028	.031
	Exper.	-.001	.005	-.000	.008
4	Total	.031	.041	.033	.044
	Talent	.032	.035	.033	.035
	Exper.	-.001	.006	.001	.010
5-6	Total	.038	.048	.040	.053
	Talent	.039	.041	.039	.042
	Exper.	-.001	.007	.001	.011
7-10	Total	.039	.050	.041	.052
	Talent	.039	.042	.039	.041
	Exper.	.000	.008	.002	.012
11+	Total	.042	.053	.044	.054
	Talent	.039	.044	.039	.039
	Exper.	.003	.009	.005	.015

Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled “Actual”) or a randomly selected feasible allocation (columns labeled “Random”) among all school-year-field combinations with the number of eligible teachers specified by the row label. Classroom-level gains are pooled across the three fields (math, science, and social studies). The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the components of the gains attributable to task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t - 1$ prior to simulated reassignment in year t . *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t - 1$. See Section 3.7.1 and Appendix Section F for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers’ task-specific talent are derived from the Full Specification (equation (3.13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher’s task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Figure 3.1: Fraction of Teachers Starting New or Discontinuing Existing Courses by Year of General Experience



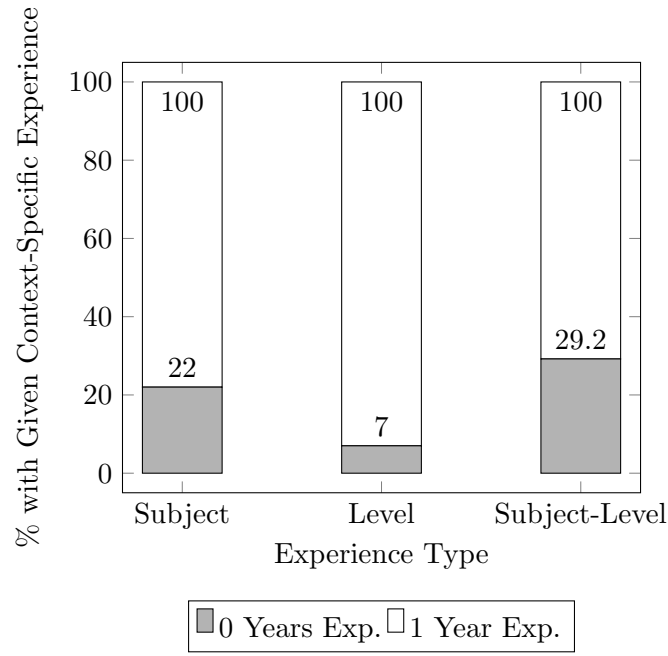
(a) First Time Teaching Course



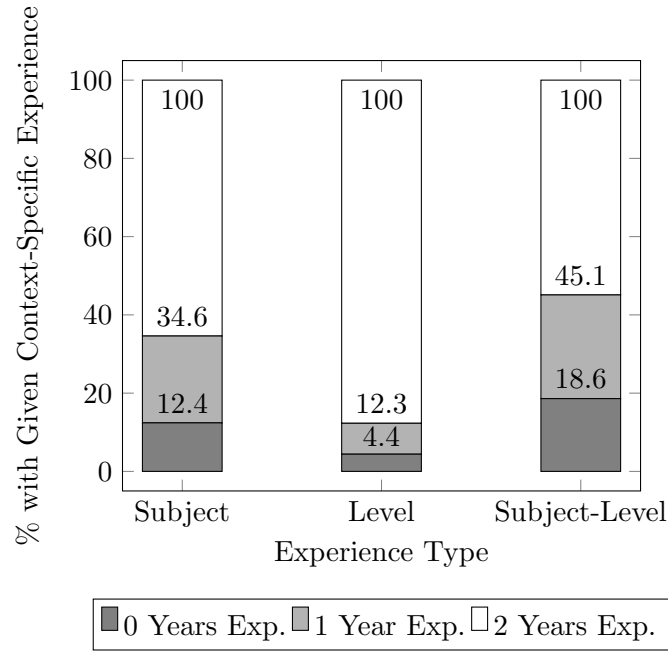
(b) Last Time Teaching Course

Notes: Panel A plots the fraction of teachers with the given number of years of general experience that teach a new subject, level, and subject-level combination, respectively, in that year that they have not previously taught. Panel B plots the fraction of teachers with the given number of years of general experience that discontinue teaching at least one subject, level, and subject-level combination, respectively, after the chosen year. Counts for the number of school-teacher-year observations associated with each general experience level are: 0 (5,294), 1 (4,249), 2 (3,545), 3 (2,901), 4 (2,322), 5 (1,792), 6 (1,385), 7 (1,106).

Figure 3.2: The Distribution of Context-Specific Experience among Second- and Third-Year Teachers



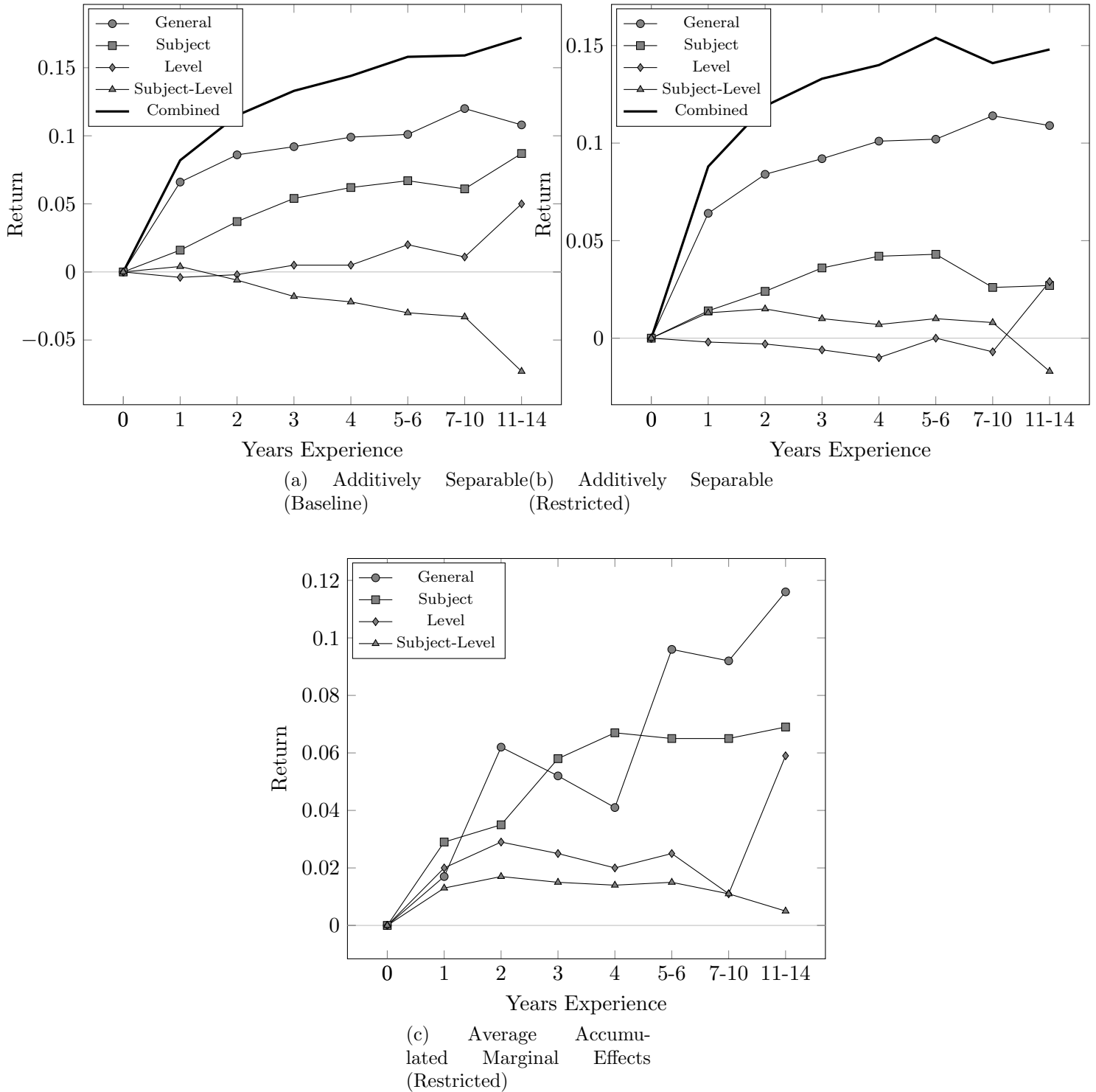
(a) Second-Year Teachers



(b) Third-Year Teachers

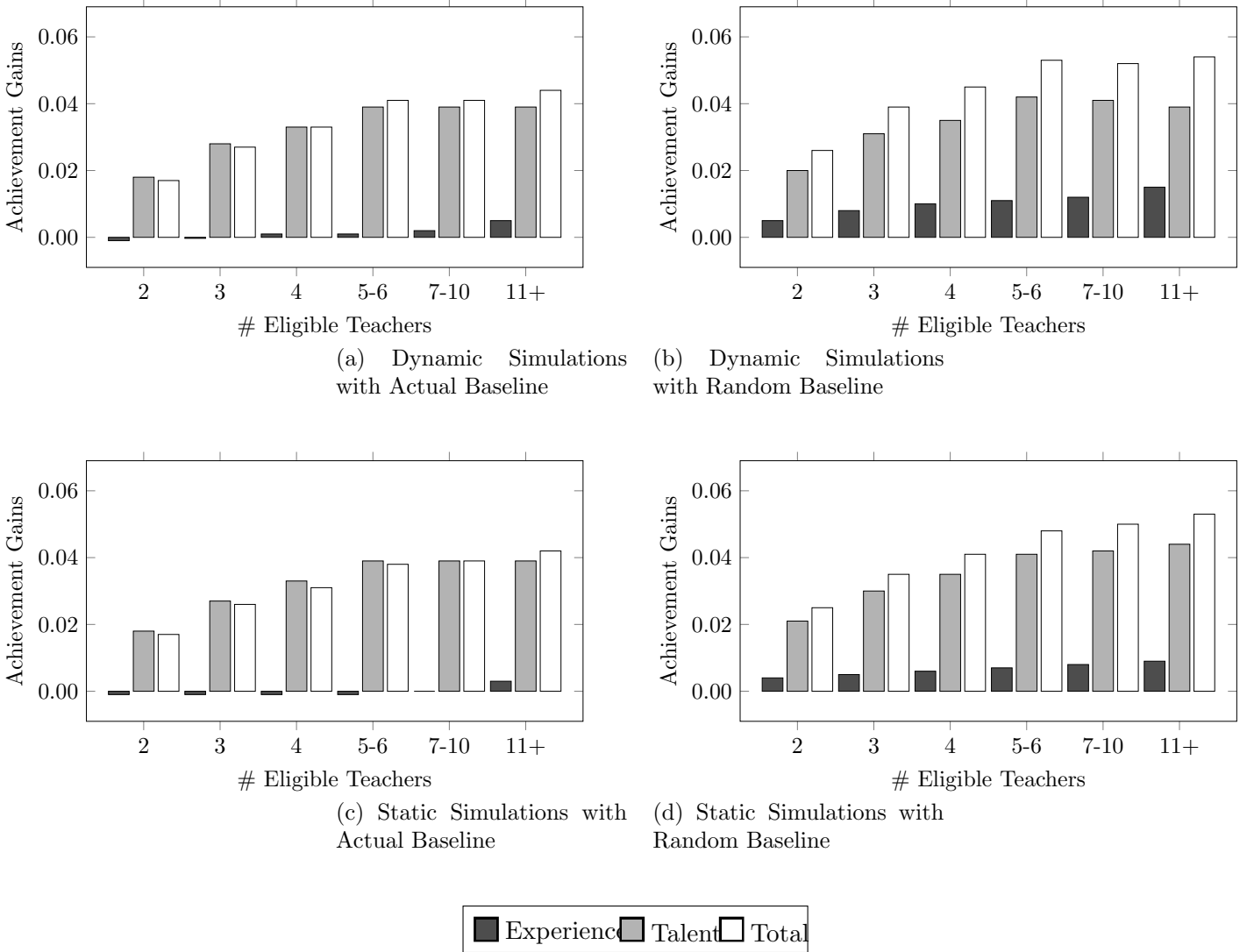
Notes: The figure displays the classroom-weighted distribution of four-dimensional experience stocks among second- and third-year teachers in our final sample. The sample includes 10,270 and 8,665 total classes taught by a second-year and third-year teacher respectively. Panel A displays the fractions of classrooms taught by second-year teachers in which the teacher has 0 versus 1 prior years of the relevant subject-, level-, and subject-level-specific experience, respectively. Panel B displays the fractions of classrooms taught by third-year teachers in which the teacher has 0, 1, and 2 prior years of the relevant subject-, level-, and subject-level-specific experience, respectively. Note that multiple subject-level combinations can be taught in a year. The full joint distribution of four dimensional experience profiles for second- and third-year teachers can be found in Appendix table F.2.

Figure 3.3: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Various Specifications)



Notes: Figures 3.3a, 3.3b, and 3.3c plot the entries from Tables 3.5, 3.6, and F.20, respectively. Refer to the notes from these tables for further detail concerning these specifications.

Figure 3.4: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)



Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in sub-figures labeled “Actual”) or a randomly selected feasible allocation (sub-figures labeled “Random”) among all school-year-field combinations with the number of eligible teachers specified on the x-axis. The gains reported are averages of classroom-level gains across all classrooms in math, science, and social studies from the final 5 years of simulated allocations (2005-2009). The white cells display the total achievement gains, while the grey and black cells display the components of the gains attributable to more efficient use of task-specific talent and task-specific experience, respectively. Subfigures labeled *Static* refer to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t - 1$ prior to simulated reassignment in year t . Subfigures labeled *Dynamic* refer to simulations in which simulated classroom assignments from 1995 through year $t - 1$ are used construct the teacher experience stocks that determine the simulated reassignment in year t . See Section 3.7.1 and Appendix Section F for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers’ task-specific talent are derived from the Full Specification (equation (3.13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher’s task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Bibliography

- Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and Student Achievement in Chicago Public Schools. *Journal of Labor Economics* 25(1), 95–135.
- Abdulkadiroglu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. a. Pathak (2011, jul). Accountability and flexibility in public schools: Evidence from boston’s charters and pilots. *Quarterly Journal of Economics* 126(2), 699–748.
- Abulkadiroglu, A., J. D. Angrist, and P. Pathak (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica* 82(1), 137–196.
- Altonji, J. G. and R. K. Mansfield (2014). Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects. Technical report, National Bureau of Economic Research.
- Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103(484), 1481–1495.
- Andrews, R., S. Imberman, and M. Lovenheim (2016). Recruiting and Supporting Low-Income, High-Achieving Students At Flagship Universities. *NBER Working Paper Series* 22260, 1–58.
- Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016). Stand and deliver: Effects of Boston’s charter high schools on college preparation, entry, and choice. *Journal of Labor Economics* 34(2).
- Angrist, J. D., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics* 1(1), 136–163.
- Angrist, J. D. and K. Lang (2004, dec). Does school integration generate peer effects? Evidence from Boston’s metco program. *American Economic Review* 94(5), 1613–1634.
- Angrist, J. D. and V. Lavy (1999). Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics* 114(2), 533–575.
- Angrist, J. D. and V. Lavy (2009). The Effects of High Stakes High School Achievement Awards : Evidence from a Randomized Trial. *The American Economic Review* 99(4), 1384–1414.
- Arsen, D. and Y. Ni (2012a). Resource Allocation in Charter and Traditional Public Schools: Is Administration Less Costly in Charter Schools. *Education Policy Analysis Archives*, 0–34.
- Arsen, D. and Y. Ni (2012b). The Effects of Charter School Competition on School District Resource Allocation. *Educational Administration Quarterly* 48(1), 3–38.
- Atteberry, A., S. Loeb, and J. Wyckoff (2013). Do first impressions matter? improvement in early career teacher effectiveness. Technical report, National Bureau of Economic Research.
- Aucejo, E. (2011). Assessing the role of teacher and student interactions. Working paper.
- Bifulco, R., C. D. Cobb, and C. Bell (2009, dec). Can Interdistrict Choice Boost Student Achievement? The Case of Connecticut’s Interdistrict Magnet School Program. *Educational Evaluation and Policy Analysis* 31(4), 323–345.
- Bifulco, R. and R. Reback (2014). Fiscal Impacts of Charter Schools: Lessons from New York. *Education Finance and Policy* 9, 86–107.
- Billings, S. B., D. J. Deming, and J. Rockoff (2014). School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg. *Quarterly Journal of Economics* 129, 435–476.
- Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics* 114(2), 577–599.
- Boyd, D., H. Lankford, S. Loeb, J. Rockoff, and J. Wyckoff (2008). The narrowing gap in new york city teacher qualifications

- and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management* 27(4), 793–818.
- Buerger, C. (2014). Unintended Effects of Charter School Programs.
- Card, D. (1990). Unexpected Inflation, Real Wages, and Employment Determination in Union Contracts. *The American Economic Review* 80(4), 669–688.
- Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone’s kids. *American Economic Journal: Applied Economics* 2(1), 211–228.
- Cascio, E. and A. Narayan (2015). Who Needs a Fracking Education? The Educational Response to Low-Skill Biased Technological Change.
- Cellini, S. R., F. Ferreira, and J. Rothstein (2010). The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design. *Quarterly Journal of Economics* 125(1), 215–261.
- Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics* 110, 124–146.
- Chakrabarti, R. and J. Roy (2016). Do Charter Schools Crowd Out Private School Enrollment? Evidence from Michigan. *Journal of Urban Economics* 91, 88–103.
- Chetty, R., J. N. Friedman, and J. Rockoff (2016). Using lagged outcomes to evaluate bias in value-added models. Nber working paper 21961, National Bureau of Economic Research.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *The American Economic Review* 104(9), 2593–2632.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review* 104(9), 2633–2679.
- Clement, M. B., L. Koonce, and T. J. Lopez (2007). The roles of task-specific forecasting experience and innate ability in understanding analyst forecasting performance. *Journal of Accounting and Economics* 44(3), 378–398.
- Clotfelter, C., H. Ladd, and J. Vigdor (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics of Education Review* 26(6), 673–682.
- Clotfelter, C., H. Ladd, and J. Vigdor (2008). School Segregation Under Color-Blind Jurisprudence: The Case of North Carolina. *Virginia Journal of Social Policy and the Law* 16(Fall), 46–86.
- Clotfelter, C., J. Vigdor, and H. Ladd (2006). Federal oversight, local control, and the specter of “resegregation” in Southern schools. *American Law and Economics Review* 8(2), 347–389.
- Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41(4), 778–820.
- Condie, S., L. Lefgren, and D. Sims (2014). Teacher heterogeneity, value-added and education policy. *Economics of Education Review* 40, 76–92.
- Craig, S. G., S. Imberman, and A. Perdue (2013). Does it pay to get an A? School resource allocations in response to accountability ratings. *Journal of Urban Economics* 73(1), 30–42.
- Crain, R. L., A. L. Heebner, and Y.-p. Si (1992). The Effectiveness of New York City’s Career Magnet Schools: An Evaluation of Ninth-Grade Performance Using an Experimental Design. Technical report, National Center for Research in Vocational Education, Berkeley, CA.
- Crain, R. L. and R. Thaler (1999). The Effects of Academic Career Magnet Education on High Schools and Their Graduates. Technical report, National Center for Research in Vocational Education, Berkeley, CA.
- Cullen, J. and B. Jacob (2007). Is gaining access to selective elementary schools gaining ground? Evidence from randomized lotteries. *The Problems of Disadvantaged Youth: An ...* (October), 43–84.
- Cullen, J. B., B. a. Jacob, and S. Levitt (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica* 74(5), 1191–1230.
- DeAngelo, G. and E. Owens (2012). Learning the Ropes: Task-Specific Experience and the Output of Idaho State Troopers. Working paper.
- Dee, T. and X. Lan (2015). The achievement and course-taking effects of magnet schools: Regression-discontinuity evidence from urban China. *Economics of Education Review* 47, 128–142.
- Dee, T. S. and H. Fu (2004). Do charter schools skim students or drain resources? *Economics of Education Review* 23(3), 259–271.
- Deming, D. J. (2011). Better schools, less crime? *Quarterly Journal of Economics* 126(4), 2063–2115.
- Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014, jan). School choice, school quality, and postsecondary

- attainment. *American Economic Review* 104(3), 991–1013.
- Dobbie, W. and R. G. Fryer (2014, jul). The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics* 6(3), 58–75.
- Duflo, E., P. Dupas, and M. Kremer (2011). Peer Effects, Teacher Incentives and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review* 101(5), 1739–1774.
- Dynarski, S., S. Hemelt, and J. Hyman (2013). The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes. *NBER Working Paper 19552*, 1–53.
- Ellison, G. and P. A. Pathak (2016). The Efficiency of Race-Neutral Alternatives to Race-Based Affirmative Action: Evidence from Chicago’s Exam Schools. *NBER Working Paper Series 22589*, 1–60.
- Engberg, J., D. Epple, J. Imbrogno, H. Sieg, and R. Zimmer (2014). Evaluating Education Programs That Have Lotteried Admission and Selective Attrition. *Journal of Labor Economics* 32(1), 27–63.
- Epple, D., R. Romano, and R. Zimmer (2015). Charter Schools: A Survey of Research on their Characteristics and Effectiveness.
- Friedman, M. (1955). *The role of government in education*. Rutgers University Press.
- Friedman, M. (1997). Public schools: Make them private. *Education Economics* 5(3), 341–344.
- Gamoran, A. and B. P. An (2016). Effects of School Segregation and School Resources in a Changing Policy Context. *Educational Evaluation and Policy Analysis* 38(1), 43–64.
- GAO (2016). Better Use of Information Could Help Agencies Identify Disparities and Address Racial Discrimination. Technical Report April, United States Government Accountability Office, Washington DC.
- Garlick, R. and J. Hyman (2015). Data vs Methods : Quasi-Experimental Evaluation of Alternative Sample Selection Corrections for Missing College Entrance Exam Score Data 1.
- Gathmann, C. and U. Schoenberg (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics* 28(1), 1–49.
- Gibbons, R. and M. Waldman (2004). Task-specific human capital. *The American Economic Review* 94(2), 203–207.
- Goldhaber, D. and D. D. Chaplin (2015). Assessing the rothstein falsification test: Does it really show teacher value-added models are biased? *Journal of Research on Educational Effectiveness* 8(1), 8–34.
- Guryan, J. (2004). Desegregation and Black Dropout Rates. *American Economic Review* 94(4), 919–943.
- Ham, J., X. Li, and L. Shore-Sheppard (2009). Seam Bias, Multiple-State, Multiple-Spell Duration Models and the Employment Dynamics of Disadvantaged Women. *NBER Working Paper Series #15151*.
- Hanushek, E., J. Kain, D. O’Brien, and S. Rivkin (2005). The Market for Teacher Quality. Nber working paper 11154, National Bureau of Economic Research, Inc.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal* 113(485), F64–F98.
- Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics* 27(3), 349–383.
- Hanushek, E. A., J. F. Kain, S. G. Rivkin, and G. F. Branch (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics* 91(5-6), 823–848.
- Hanushek, E. A. and S. G. Rivkin (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of Policy Analysis and Management* 28(3), 366–393.
- Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? an examination of the statistical properties and policy alternatives. *Education Finance and Policy* 4(4), 319–350.
- Harris, D. N. and T. R. Sass (2011). Teacher training, teacher quality and student achievement. *Journal of public economics* 95(7), 798–812.
- Hastings, J., C. Neilson, and S. S. Zimmerman (2012). The effect of school choice on intrinsic motivation and academic outcomes. *NBER Working Paper No.18324* 18324.
- Hastings, J. S., T. J. Kane, and D. O. Staiger (2006). Gender and performance: Evidence from school assignment by randomized lottery. *American Economic Review* 96(2), 232–236.
- Hastings, J. S., T. J. Kane, and D. O. Staiger (2009). Heterogeneous Preferences and the Efficacy of Public School Choice. *NBER Working Paper 2145*, 1–46.
- Hastings, J. S. and J. M. Weinstein (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics* 123(4), 1373–1414.
- Hong, K. and R. Zimmer (2016). Does Investing in School Capital Infrastructure Improve Student Achievement?

- Hoxby, C. M. (1996). How Teachers' Unions Affect Education Production. *The Quarterly Journal of Economics* 111(3), 671–718.
- Hoxby, C. M. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Paper Series No. 7867*, 64.
- Hoxby, C. M. (2002). Would School Choice Change the Teaching Profession? *The Journal of Human Resources* 37(4), 846–891.
- Hoxby, C. M. (2003a). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review* 10, 9–65.
- Hoxby, C. M. (2003b). School choice and school productivity: Could school choice be a tide that lifts all boats? In *The Economics of School Choice*, pp. 289–342.
- Hoxby, C. M. (2006). The Supply of Charter Schools. In P. T. Hill (Ed.), *Education Next*, Volume 6, Chapter 1, pp. 15–44. Hoover Institution Press.
- Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Imberman, S. (2011). The effect of charter schools on achievement and behavior of public school students. *Journal of Public Economics* 95(7-8), 850–863.
- Imberman, S., M. Rourke, and M. Naretta (2016). Capitalization of Charter Schools into Residential Property Values. *Education Finance and Policy*, forthcoming.
- Jackson, C. K. (2013a). Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *Review of Economics and Statistics* 95, 1096–1113.
- Jackson, C. K. (2014). Do High School Teachers Really Matter? *Journal of Labor Economics* 32(4).
- Jackson, C. K. and E. Bruegmann (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1(4), 85–108.
- Jackson, K. (2009). Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation. *Journal of Labor Economics* 27(2), 213–256.
- Jackson, K. (2012). School competition and teacher labor markets: Evidence from charter school entry in North Carolina. *Journal of Public Economics* 96, 431–448.
- Jackson, K. (2013b, dec). Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. *Journal of Public Economics* 108, 63–77.
- Jacob, B. and J. Rockoff (2011). *Organizing schools to improve student achievement: start times, grade configurations, and teacher assignments*. Brookings Institution, Hamilton Project.
- Johnson, R. C. (2015). Long-Run Impacts of School Desegregation and School Quality on Adult Attainment. *NBER Working Paper 16664*, 1–87.
- Kinsler, J. (2012). Assessing rothstein's critique of teacher value-added models. *Quantitative Economics* 3(2), 333–362.
- Lazear, E. P. (2001). Educational Production. *Quarterly Journal of Economics* 116(3), 777–803.
- Linick, M. (2014). Measuring Competition: Inconsistent Definitions, Inconsistent Results. *Education Policy Analysis Archives* 22(16), 1–23.
- Lockwood, J. and D. McCaffrey (2009). Exploring Student-Teacher Interactions in Longitudinal Achievement Data. *Education Finance and Policy* 4(4), 439–467.
- Lovenheim, M. and P. Walsh (2014). Does Choice Increase Information? Evidence from Online School Search Behavior.
- LSC (2011). School Funding Complete Resource. Technical report, Legislative Service Commission, Columbus, OH.
- Lutz, B. (2011). The end of court-ordered desegregation. *American Economic Journal: Economic Policy* 3(May), 130–168.
- Mansfield, R. (Forthcoming). Teacher Quality and Student Inequality. *Journal of Labor Economics*.
- Martorell, P., K. Stange, and I. McFarlin (2015). Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement.
- Massey, D. S. and N. A. Denton (1988). The Dimensions of Residential Segregation. *Social Forces* 67(2), 281–315.
- Master, B., S. Loeb, C. Whitney, and J. Wyckoff (2012). Different skills: Identifying differentially effective teachers of english language learners. Working paper.
- Meyer, B. D. and J. X. Sullivan (2008). Changes in the consumption, income, and well-being of single mother headed families. *American Economic Review* 98(5), 2221–2241.
- ODE (2011). Guide To Understanding Ohio's Accountability System: 2010-2011. Technical report, Ohio Department of

Education.

- ODE (2014). Public Charter/Community Schools: Guidance for New Developers. Technical Report June, Ohio Department of Education, Columbus, OH.
- O'Donnell, P. (2014, jul). Ohio is the "Wild, Wild West" of charter schools, says national group promoting charter standards.
- Okui, R. (2014). Asymptotically unbiased estimation of autocovariances and autocorrelations with panel data in the presence of individual and time effects. *Journal of Time Series Econometrics* 6(2), 129–181.
- Orfield, G., J. Ee, E. Frankenberg, and G. Siegel-Hawley (2016). Brown At 62: School Segregation By Race, Poverty and State. Technical report, Civil Rights Project, UCLA, Los Angeles.
- Ost, B. (2014). How Do Teachers Improve? The Relative Importance of Specific and General Human Capital. *American Economic Journal: Applied Economics* 6(2), 127–51.
- Ost, B. and J. C. Schiman (2015). Grade-specific experience, grade reassignments, and teacher turnover. *Economics of Education Review* 46, 112–126.
- Papay, J. P. and M. A. Kraft (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*.
- Park, A., X. Shi, C. tai Hsieh, and X. An (2015). Magnet high schools and academic performance in China: A regression discontinuity design. *Journal of Comparative Economics* 43(4), 825–843.
- Pei, Z. (2015). Eligibility Recertification and Dynamic Opt-in Incentives in Income-tested Social Programs: Evidence from Medicaid / CHIP.
- Pischke, J.-S. (1995). Individual Income, Incomplete Information, and Aggregate Consumption. *Econometrica* 63(4), 805–840.
- Polataev, M. and C. Robinson (2008). Teachers, Schools, and Academic Achievement. *Journal of Labor Economics* 26(3), 387–420.
- Reardon, S. F., E. T. Grewal, D. Kalogrides, and E. Greenberg (2012). Brown Fades: The End of Court-Ordered School Desegregation and the Resegregation of American Public Schools. *Journal of Policy Analysis and Management* 31(4), 876–904.
- Reber, S. J. (2010). School Desegregation and Educational Attainment for Blacks. *The Journal of Human Resources* 45(4), 893–914.
- Rivkin, S., E. Hanushek, and J. Kain (2005). Teachers, Schools, and Academic Achievement. *Econometrica* 73(2), 417–458.
- Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review: Papers and Proceedings of the One Hundred Sixteenth Annual Meeting of the American Economic Association* 94(2), 247–252.
- Rossell, C. (2003). The Desegregation Efficiency of Magnet Schools. *Urban Affairs Review* 38(May), 697–725.
- Rossell, C. (2005). Magnet Schools. *Education Next* 5(2).
- Rossell, C. H. and D. J. Armor (1996). The Effectiveness of School Desegregation Plans, 1968-1991. *American Politics Research* 24, 267–302.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics* 125(1), 175–214.
- Sacerdote, B. (2011). Peer Effects in Education: How might they work, how big are they and how much do we know Thus Far? *Handbook of the Economics of Education* 3(11), 249–277.
- Sass, T. R., A. Semykina, and D. N. Harris (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review* 38, 9–23.
- Sullivan, M. and M. Sobul (2010). Property Taxation and School Funding. Technical Report February, Ohio Department of Taxation.
- Sullivan, M. D., D. B. Campbell, and B. Kisida (2008). The Muzzled Dog That Didn't Bark: Charters and the Behavioral Response of D.C. Public Schools. *School Choice Demonstration Project*, 1–43.
- Taylor, L. L. (2006). The Labor Market Impact of School Choice: Charter Competition and Teacher Compensation. *Advances in Applied Microeconomics* 14, 257–279.
- Taylor, L. L. (2010). Competition and teacher pay. *Economic Inquiry* 48(3), 603–620.
- Vedder, R. and J. Hall (2000). Private school competition and public school teacher salaries. *Journal of Labor Research* 21(1), 161–168.
- Vigdor, J. and J. Ludwig (2008). Segregation and the Test Score Gap. *Steady Gains and Stalled Progress*, 181–211.

- Vigdor, J. and T. J. Nechyba (2007). Peer Effects in North Carolina Public Schools. In P. E. Peterson and L. Woessmann (Eds.), *Schools and the equal opportunity problem*, Volume 49, pp. 73–103. MIT Press.
- Wiswall, M. (2013). The Dynamics of Teacher Quality. *Journal of Public Economics* 100, 61–78.
- Yamaguchi, S. (2012). Tasks and heterogeneous human capital. *Journal of Labor Economics* 30(1), 1–53.

Part I

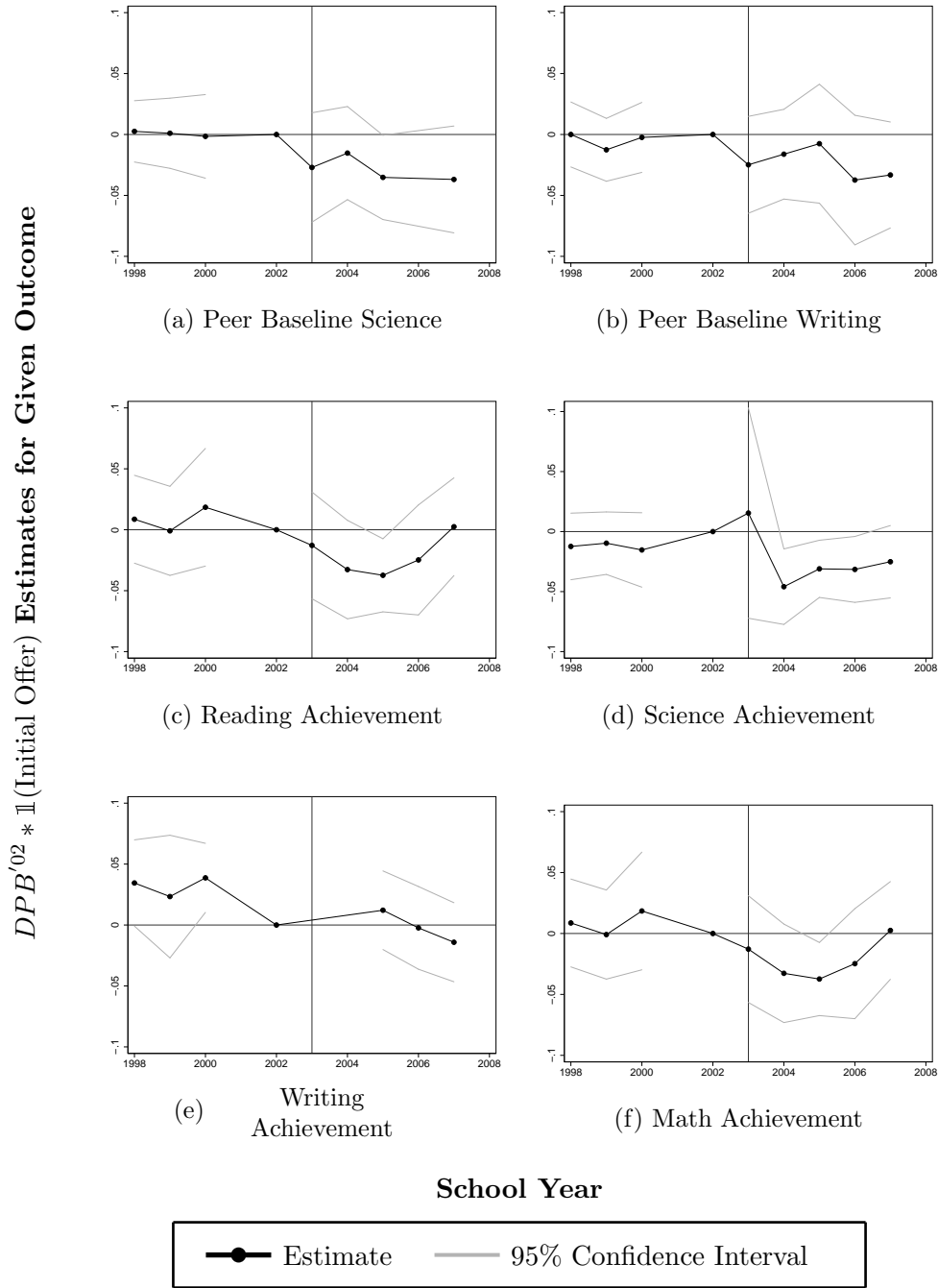
Appendices

Chapter 4

Segregation Appendix

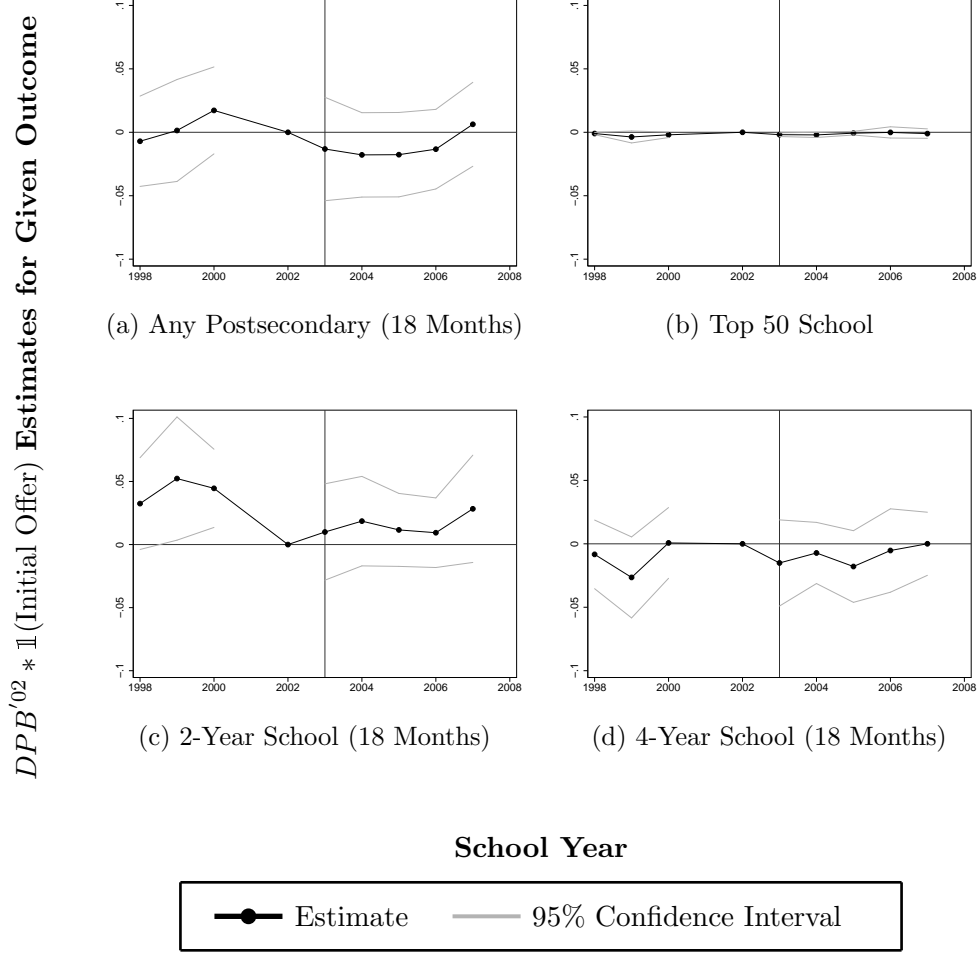
A Differential Trends Test: Main Outcomes

Figure A.1: Trends in Various Outcomes by Lottery Racial Disparity (DPB'^{02})



Notes: Each figure presents the effect of receiving an initial seat offer to a magnet school through a lottery with a 1 percentage point larger disparity between the percentage of black students in the lottery pool and the percentage of black student receiving offers in 2002 (DPB'^{02}) on the given current outcome. Regressions are estimated using (1.5) as explained in Section 1.6.1. Each regression is respectively run using the sample restrictions for the given outcome in the footnotes of Table 1.8. The reference line in 2003 denotes the first year the LUSD implemented the consolidated lottery system.

Figure A.2: Trends in Various Outcomes by Lottery Racial Disparity (DPB'^{02})



Notes: Each figure presents the effect of receiving an initial seat offer to a magnet school through a lottery with a 1 percentage point larger disparity between the percentage of black students in the lottery pool and the percentage of black student receiving offers in 2002 (DPB'^{02}) on the given current outcome. Regressions are estimated using (1.5) as explained in Section 1.6.1. Each regression is respectively run using the sample restrictions for the given outcome in the footnotes of Table 1.9. The reference line in 2003 denotes the first year the LUSD implemented the consolidated lottery system.

B Disaggregated Magnet Effects

B.1 Effects of Magnet Enrollment on Academic Achievement

Table B.1 presents the effect of magnet enrollment on subject-specific statewide achievement.¹ Panel A provides estimates for the effect of magnet enrollment on subsequent achievement among the pooled regression sample.² I am unable to detect statistically significant differences between magnet and traditional enrollees.

Panel B disaggregates these estimates among sub-groups by student race, gender, and baseline aptitude. Specifically, I calculate whether each student's baseline math achievement is above or below the district's median for the given year. Gains to middle school science achievement are local to non-black and female students. One possible explanation for why subject-specific effects can differ so drastically by student subgroups is that magnet schools across the district vary considerably in their teaching strategies and specialties. Particularly effective schools could be driving estimates for certain student subgroups. However, because the focus of this paper is estimating segregation effects, I leave this particular exploration to future work.

These estimates show that the gains to magnet enrollment over traditional school enrollment are generally small and statistically insignificant. Because these schools are trying to attract non-black enrollment, it is sensible that the largest magnet school gains are concentrated among non-black students. Due to the localized nature of the returns to magnet enrollment, I conclude that enrolling in a magnet or traditional school in this LUSD yields similar achievement returns.

¹The grades that achievement tests were administered slightly changed over time. Middle school tests were administered during 8th grade in 1995, 2003, and 2004 and in 6th grade all other years.

²First-stage estimates, observation counts, weak F tests, and outcome means are presented in tables B.2, B.3, B.4, and B.5 respectively.

Table B.1: Effect of Magnet Enrollment on Achievement – Application Level

	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>Pooled Sample</i>				
Enrolled in Magnet	0.038 (0.078)	0.161 (0.145)	0.171 (0.123)	-0.032 (0.123)
<i>First Stage:</i> Initial Offer	0.193*** (0.069)	0.193*** (0.069)	0.195*** (0.069)	0.199*** (0.073)
Observations	12,198	12,170	10,587	9,463
Panel B: <i>Effects by Subgroup</i>				
Non-Black	0.076 (0.109)	0.082 (0.106)	0.422*** (0.152)	0.022 (0.124)
Black	0.006 (0.084)	0.178 (0.180)	0.082 (0.136)	-0.046 (0.142)
Male	-0.037 (0.132)	0.155 (0.233)	0.162 (0.198)	0.123 (0.146)
Female	0.095 (0.068)	0.167 (0.113)	0.207** (0.097)	-0.163 (0.157)
Above Median Baseline Math Score	0.025 (0.091)	0.169 (0.196)	0.366 (0.231)	-0.033 (0.160)
Below Median Baseline Math Score	-0.058 (0.144)	0.089 (0.187)	-0.004 (0.195)	-0.128 (0.220)
Black, Female	0.050 (0.086)	0.211 (0.164)	0.094 (0.124)	-0.158 (0.218)
Black, Male	-0.040 (0.142)	0.152 (0.241)	0.109 (0.225)	0.087 (0.157)
Black, Above Median	-0.001 (0.116)	0.300 (0.278)	0.303 (0.308)	0.007 (0.190)
Black, Below Median	-0.093 (0.166)	0.090 (0.213)	-0.056 (0.212)	-0.182 (0.267)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, year-of-test and risk-sets. I instrument for endogenous magnet attendance variable with whether the student received an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions i.e., the student must have applied to a magnet school lottery in 6th grade between 1998 and 2007, must be in their first year attending the grade of the lottery application (no grade retention), and must not be eligible for special education. The regressions in this table further condition on having non-missing outcome information. Regressions are weighted by one over the number of lotteries applied to by the student in the given year. The regressions in this table further condition on having non-missing outcome information. First-stage estimates, observation counts, weak IV tests, and outcome means are provided in tables B.2, B.3, B.4, and B.5, respectively.

Table B.2: First-Stage for Table B.1

	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>Pooled Sample</i>				
<i>First Stage: Initial Offer</i>	0.193*** (0.069)	0.193*** (0.069)	0.195*** (0.069)	0.199*** (0.073)
Panel B: <i>Effects by Subgroup</i>				
Non-Black	0.318*** (0.090)	0.311*** (0.089)	0.348*** (0.093)	0.334*** (0.097)
Black	0.171*** (0.066)	0.172*** (0.066)	0.170*** (0.066)	0.177** (0.070)
Male	0.183*** (0.063)	0.180*** (0.063)	0.188*** (0.062)	0.199*** (0.070)
Female	0.201*** (0.073)	0.201*** (0.074)	0.197** (0.077)	0.195** (0.077)
Above Median Baseline Math Score	0.188*** (0.071)	0.188*** (0.071)	0.187** (0.076)	0.196** (0.083)
Below Median Baseline Math Score	0.174** (0.074)	0.174** (0.074)	0.178** (0.075)	0.174** (0.079)
Black, Female	0.173** (0.069)	0.175** (0.069)	0.164** (0.070)	0.167** (0.071)
Black, Male	0.168*** (0.063)	0.168*** (0.063)	0.174*** (0.062)	0.184*** (0.070)
Black, Above Median	0.159** (0.067)	0.163** (0.068)	0.155** (0.069)	0.164** (0.075)
Black, Below Median	0.154** (0.072)	0.154** (0.071)	0.156** (0.073)	0.153** (0.078)

Notes: This Table provides first-stage estimates for Table B.1. See notes in Table B.1 for specification details.

Table B.3: Observation Counts for Table B.1

	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>Pooled Sample</i>				
Enrolled in Magnet	12,198	12,170	10,587	9,463
Panel B: <i>Effects by Subgroup</i>				
Non-Black	1,891	1,864	1,625	1,494
Black	10,307	10,306	8,962	7,969
Male	5,369	5,349	4,757	4,136
Female	6,829	6,821	5,830	5,327
Above Median Baseline Math Score	5,723	5,689	4,922	4,313
Below Median Baseline Math Score	4,671	4,682	4,147	3,532
Black, Female	5,785	5,784	4,950	4,487
Black, Male	4,522	4,522	4,012	3,482
Black, Above Median	4,446	4,439	3,831	3,328
Black, Below Median	4,313	4,321	3,816	3,248

Notes: This Table provides observation counts for Table B.1. See notes in Table B.1 for specification details.

Table B.4: Weak Instruments F Statistics for Table B.1

	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>Pooled Sample</i>				
Enrolled in Magnet	7.945	7.893	7.906	7.436
Panel B: <i>Effects by Subgroup</i>				
Non-Black	12.472	12.167	14.131	11.884
Black	6.764	6.771	6.676	6.436
Male	8.310	8.199	9.196	8.192
Female	7.491	7.492	6.635	6.390
Above Median Baseline Math Score	6.995	6.976	6.139	5.532
Below Median Baseline Math Score	5.513	5.536	5.665	4.834
Black, Female	6.356	6.388	5.521	5.585
Black, Male	7.213	7.211	7.918	6.998
Black, Above Median	5.667	5.781	5.009	4.733
Black, Below Median	4.621	4.642	4.541	3.884

Notes: Table reports weak instrument tests using the Kleibergen-Paap Wald F statistic for Table B.1. See notes in Table B.1 for specification details.

Table B.5: Control Group Outcome Means for Table B.1

	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>Pooled Sample</i>				
Enrolled in Magnet	0.326	0.169	0.082	0.337
Panel B: <i>Effects by Subgroup</i>				
Non-Black	0.729	0.731	0.585	0.622
Black	0.270	0.092	0.018	0.301
Male	0.230	0.131	0.104	0.133
Female	0.399	0.199	0.064	0.490
Above Median Baseline Math Score	0.617	0.571	0.372	0.616
Below Median Baseline Math Score	-0.063	-0.370	-0.290	0.059
Black, Female	0.349	0.122	0.009	0.454
Black, Male	0.166	0.053	0.029	0.102
Black, Above Median	0.564	0.489	0.306	0.584
Black, Below Median	-0.075	-0.387	-0.306	0.061

Notes: This Table provides outcome means among students not offered a magnet seat for Table B.1. See notes in Table B.1 for specification details.

B.2 Effects of Magnet Enrollment on ACT Testing

Estimates of the effect of magnet enrollment on ACT test taking behavior and composite scores are provided in Panel A of Table B.6 for the pooled sample and Panel B for subgroup analyses.³ In column 1, I estimate that enrollment in a magnet middle school has no effect on ACT test taking. Columns 2 presents effects on composite ACT scores. Magnet middle schools appear to decrease composite scores by about half of a point, though I cannot rule out change findings. Subgroup estimates are again imprecise, but the possibility of heterogeneous impacts. Magnet schools increase ACT test taking for non-black, low-aptitude, and male students, while they decrease test taking for females.⁴ If magnet middle schools increase the proportion of students taking the ACT it is likely that students at the lower end of the achievement distribution are the ones induced to take the test.

In Appendix B.2, I show how the simulation procedure used by Garlick and Hyman (2015) can be adapted to correct for selection in settings where the treatment directly affects whether the outcome of interest is observed. This method is similar to that used by Meyer and Sullivan (2008) to impute missing housing and vehicle prices. As an overview, I regress ACT scores on student demographics, prior test scores and current-school-by-year fixed effects among ACT test takers and generate a vector of residuals. Then for non-ACT takers, I predict their ACT score based on student demographics, prior math and reading test scores, and current-school-by-year fixed effects and fill in with randomly chosen errors from the first step.⁵ I then run equation (1.1) on this new outcome, repeat the procedure 50 times and report the mean of both the estimates and the standard errors.

Column 3 presents these selection-corrected estimates. As expected, magnet middle school effects are generally larger in magnitude after accounting for the change in the composition of the pool of students taking the ACT. Table B.11 presents estimates for subject-specific, selection-corrected ACT scores. Black male students improve their ACT English scores, however, non-black students perform worse on the ACT Math section.

Selection Correction

The selection-correction method is as follows:

1. Estimate the relationship between covariates and ACT scores among test takers using prior test scores and current-school-by-year fixed effects
2. Predict residuals
3. For all missing values fill in with the predicted ACT score plus a randomly drawn error term
4. Repeat K times
5. Report mean of beta vector and standard error vector

This is done so that the distribution of predicted ACT scores of is comparable to the latent distribution. Not incorporating errors would understate the variance of predicted ACT scores (Garlick and Hyman, 2015). This procedure introduces measurement error for non-test takers, but because this is measurement error in the outcome, it will increase noise, but will not necessarily induce a estimation

bias.

³For the small share of students taking the SAT instead of the ACT, I convert scores to their ACT counterparts.

⁴While each of these suggestive estimates are not significantly different than zero, many are larger than their standard errors.

⁵I also predict scores for student for whom I have no ACT test taking information. Thus, sample sizes for selection-corrected outcomes are larger than the original ACT test taking outcome.

Table B.6: Effect of Magnet Enrollment on ACT Outcomes – Application Level

	Middle School		
	Took ACT (1)	Composite Score (2)	Selection- Corrected Composite (3)
Panel A: <i>Pooled Sample</i>	-0.000 (0.069)	-0.453 (1.042)	0.172 (0.479)
Panel B: <i>Effects by Subgroup</i>			
Non-Black	0.112 (0.071)	-1.108 (2.180)	0.321 (1.096)
Black	-0.031 (0.089)	-0.513 (1.223)	0.081 (0.535)
Male	0.210 (0.133)	-1.982 (1.998)	-0.517 (0.588)
Female	-0.146 (0.113)	-0.006 (0.900)	0.239 (0.558)
Above Median Baseline Math Score	-0.045 (0.082)	-0.567 (1.273)	0.451 (0.836)
Below Median Baseline Math Score	0.197 (0.212)	0.665 (1.072)	1.362 (1.010)
Black, Male	0.249 (0.167)	-1.417 (2.494)	0.476 (0.791)
Black, Female	-0.220 (0.161)	-0.239 (0.939)	1.057 (0.908)
Black, Above Median	-0.085 (0.126)	-0.574 (1.460)	0.137 (1.013)
Black, Below Median	0.200 (0.252)	0.370 (1.307)	1.292 (1.066)

Notes: *, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively. Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions specified in the notes for Table B.1. Estimates from Panel A are only run among the sample of students with valid test scores. Estimates from Panel B correct for selection by imputing scores for non-test-takers as is explained in Appendix B.2. Analysis is run on application-level data meaning that a student with multiple applications in a given year will appear multiple times in the data. Regressions are weighted by one over the number of lotteries applied to by the student in the given year. First-stage estimates, observation counts, weak IV tests, and outcome means are provided in Tables B.7, B.8, B.10, and B.9, respectively.

Table B.7: First-Stage Estimates for Table B.6

	Middle School		
	Took ACT (1)	Composite Score (2)	Selection- Corrected Composite (3)
Panel A: <i>Pooled Sample</i>	0.170*** (0.066)	0.182*** (0.068)	0.205*** (0.074)
Panel B: <i>Effects by Subgroup</i>			
Non-Black	0.327*** (0.094)	0.277** (0.113)	0.326*** (0.089)
Black	0.148** (0.062)	0.167** (0.067)	0.184** (0.073)
Male	0.159*** (0.056)	0.133** (0.058)	0.191*** (0.066)
Female	0.179** (0.076)	0.214*** (0.080)	0.216*** (0.080)
Above Median Baseline Math Score	0.184** (0.072)	0.181** (0.075)	0.185*** (0.070)
Below Median Baseline Math Score	0.129** (0.065)	0.164** (0.071)	0.175** (0.069)
Black, Male	0.138** (0.054)	0.123** (0.058)	0.166*** (0.060)
Black, Female	0.157** (0.071)	0.194** (0.076)	0.171** (0.067)
Black, Above Median	0.158** (0.068)	0.172** (0.077)	0.152** (0.065)
Black, Below Median	0.110* (0.063)	0.148** (0.070)	0.156** (0.068)

Notes: Table provides first-stage estimates for Table B.6. See notes of Table B.6 for regression details.

Table B.8: Observation Counts for Table B.6

	Middle School		
	Took ACT	Composite Score	Selection- Corrected Composite
	(1)	(2)	(3)
Panel A: <i>Pooled Sample</i>	7,993	4,371	14,601
Panel B: <i>Effects by Subgroup</i>			
Non-Black	1,049	581	2,449
Black	6,944	3,790	12,152
Male	3,292	1,661	6,651
Female	4,701	2,710	7,950
Above Median Baseline Math Score	4,001	2,262	6,282
Below Median Baseline Math Score	2,816	1,322	5,128
Black, Male	2,852	1,420	4,989
Black, Female	4,092	2,370	6,231
Black, Above Median	3,244	1,825	4,800
Black, Below Median	2,677	1,272	4,706

Notes: Table provides observation counts for Table B.6. See notes of Table B.6 for regression details.

Table B.9: Outcome Means for Table B.6

	Middle School	
	Took ACT	Composite Score
	(1)	(2)
Panel A: <i>Pooled Sample</i>	0.686	17.695
Panel B: <i>Effects by Subgroup</i>		
Non-Black	0.635	22.321
Black	0.691	17.208
Male	0.623	17.639
Female	0.728	17.727
Above Median Baseline Math Score	0.741	18.913
Below Median Baseline Math Score	0.597	15.599
Black, Male	0.623	17.095
Black, Female	0.738	17.271
Black, Above Median	0.749	18.321
Black, Below Median	0.604	15.550

Notes: Table provides outcome means among students not offered a magnet lottery seat for Table B.6. See notes of Table B.6 for regression details.

Table B.10: Weak Instruments F Statistics for Table B.6

	Middle School	
	Took ACT (1)	Composite Score (2)
Panel A: <i>Pooled Sample</i>	6.654	7.116
Panel B: <i>Effects by Subgroup</i>		
Non-Black	12.194	5.973
Black	5.666	6.279
Male	7.897	5.219
Female	5.609	7.134
Above Median Baseline Math Score	6.558	5.854
Below Median Baseline Math Score	3.881	5.343
Black, Male	6.453	4.555
Black, Female	4.839	6.481
Black, Above Median	5.453	4.998
Black, Below Median	2.997	4.452

Notes: Table reports weak instrument tests using the Kleibergen-Paap Wald F statistic for Table B.6. See notes of Table B.6 for regression details.

Table B.11: Effect of Magnet Enrollment on ACT Selection-Corrected Subject Scores – Application Level

	Middle School			
	Reading (1)	Math (2)	Science (3)	English (4)
Panel A: <i>Pooled Sample</i>				
Enrolled in Magnet	0.491 (0.849)	0.361 (0.607)	0.759 (0.764)	1.343 (0.891)
<i>First Stage:</i> Initial Offer	0.193*** (0.067)	0.193*** (0.067)	0.193*** (0.067)	0.193*** (0.067)
Observations	13,426	13,426	13,426	13,426
Panel B: <i>Effects by Subgroup</i>				
Non-Black	3.075 (2.376)	-1.635* (0.989)	-0.331 (1.352)	1.293 (1.484)
Black	0.176 (0.994)	0.332 (0.727)	0.898 (0.865)	1.565 (1.161)
Male	-0.078 (0.866)	-0.458 (0.685)	0.622 (0.789)	1.667* (1.003)
Female	0.711 (0.991)	0.510 (0.517)	1.216 (1.164)	1.347 (0.958)
Above Median Baseline Math Score	0.331 (1.321)	-0.103 (0.684)	0.741 (0.949)	0.867 (1.168)
Below Median Baseline Math Score	1.398 (0.990)	1.120 (0.986)	1.746 (1.456)	1.364 (1.332)
Black, Male	0.693 (0.992)	0.314 (0.893)	0.837 (0.992)	2.461* (1.329)
Black, Female	0.441 (1.007)	0.712 (1.002)	1.235 (1.414)	1.390 (1.108)
Black, Above Median	-0.708 (1.821)	0.214 (1.095)	0.438 (1.130)	0.977 (1.501)
Black, Below Median	1.172 (1.019)	1.491 (1.118)	2.121 (1.570)	1.808 (1.390)

Notes: Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions specified in the notes for Table B.1. Estimates from Panel A are only run among the sample of students with valid test scores. Estimates from Panel B correct for selection by imputing scores for non-test-takers as is explained in Appendix B.2. Analysis is run on application-level data meaning that a student with multiple applications in a given year will appear multiple times in the data. Regressions are weighted by one over the number of lotteries applied to by the student in the given year.

B.3 Effects of Magnet Enrollment on Postsecondary Attainment

In Table B.12, I present the effect of magnet enrollment on various college outcomes. Specifically, I test whether attending a magnet school affects the probability of attending any postsecondary institution, or a two-year or four-year, or a “Top 50” ranked institution within 18 months after high school graduation.⁶ Again, I find imprecisely estimated magnet effects for the pooled sample and heterogeneous effects by students subgroups. In the pooled sample, I find suggestive evidence that magnet schools induce a shift in postsecondary enrollment from 4-year to 2-year institutions. This shift is driven by an increase in 2-year attendance among black male students and a decrease in 4-year attendance among black female students. Lastly, magnet enrollment has a negligible impact on the probability of enrolling in a “Top 50” ranked institution.

Finally, I descriptively investigate how magnet attendance affects college major choices. Appendix Table B.13 presents these effects for the subset of students attending college. The small sample sizes make me hesitant to assign a causal interpretation to these estimates. I find evidence that magnet school enrollment is negatively associated with students choosing to major in Fine Arts fields.

⁶ “Top 50” ranked as measured by the 2006 US & News World Report.

Table B.12: Effect of Middle School Magnet Attendance on College Enrollment – Application Level

	Enrollment within 18 Months After Graduation			
	Any (1)	Two-Year (2)	Four-Year (3)	Top 50 School (4)
Panel A: <i>Pooled Estimates</i>				
Enrolled in Magnet School	-0.049 (0.106)	0.159 (0.150)	-0.104 (0.079)	-0.014 (0.013)
Panel B: <i>Sub-group Estimates</i>				
Non-Black	-0.012 (0.165)	0.014 (0.088)	-0.037 (0.180)	-0.004 (0.023)
Black	-0.078 (0.110)	0.203 (0.200)	-0.137 (0.089)	-0.018 (0.013)
Male	0.137 (0.107)	0.254* (0.154)	0.105 (0.085)	-0.046 (0.038)
Female	-0.153 (0.124)	0.112 (0.172)	-0.241** (0.119)	0.006 (0.009)
Above Median Baseline Math Score	-0.077 (0.128)	0.250 (0.220)	-0.164 (0.109)	-0.024 (0.017)
Below Median Baseline Math Score	0.062 (0.145)	0.221 (0.203)	-0.066 (0.199)	0.005 (0.015)
Black, Male	0.153 (0.123)	0.254 (0.203)	0.142 (0.103)	-0.029 (0.028)
Black, Female	-0.210 (0.137)	0.174 (0.229)	-0.314** (0.148)	-0.009 (0.008)
Black, Above Median	-0.116 (0.125)	0.337 (0.312)	-0.246 (0.150)	-0.040 (0.026)
Black, Below Median	0.053 (0.185)	0.222 (0.251)	-0.050 (0.142)	0.006 (0.019)

Notes: Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions specified in the notes for Table B.1. Top 50 ranking is from the US News and World Report 2006 ranking. Analysis is run on application-level data meaning that a student with multiple applications in a given year will appear multiple times in the data. Regressions are weighted by one over the number of lotteries applied to by the student in the given year.

Table B.13: Effect of Magnet Attendance on College Major and Graduation – Application Level

	College Major						
	Education (1)	Fine Arts (2)	Business (3)	Social Science (4)	English/ History (5)	Other (6)	Graduated (7)
Enrolled in Magnet Middle School	-0.028 (0.089)	-0.223** (0.099)	-0.071 (0.117)	0.047 (0.127)	0.122 (0.088)	0.036 (0.092)	-0.070 (0.055)
N	699	699	699	699	699	699	5,464

Notes: Regressions follow equation (1.1) where each outcome is regressed on a indicator equal to one if the student attended a magnet school during the year following the lottery as well as indicators for student gender, race, and risk-sets. I instrument for endogenous magnet attendance variable with whether the student receiving an initial lottery offer. Standard errors are two-way clustered by student and the enrolled school after the lottery. Each regression sample is limited to baseline sample restrictions specified in the notes for Table B.1. Analysis is run on application-level data meaning that a student with multiple applications in a given year will appear multiple times in the data. Regressions are weighted by one over the number of lotteries applied to by the student in the given year.

C First-Stage Estimates, Observation Counts, and Outcome Averages

Table C.1: First-Stage Estimates for Table 1.6

	Achievement Index (1)	Postsecondary Index (2)	Total Index (3)
Panel A: 2SLS Estimates for Pooled Sample			
<i>First Stage:</i> Initial Offer	0.192*** (0.067)	0.165** (0.065)	0.193*** (0.067)
Panel B: 2SLS Estimates by Subgroup			
Non-Black	0.318*** (0.089)	0.304*** (0.101)	0.321*** (0.087)
Black	0.169*** (0.065)	0.145** (0.060)	0.169*** (0.064)
Male	0.180*** (0.061)	0.160*** (0.054)	0.180*** (0.061)
Female	0.201*** (0.073)	0.172** (0.075)	0.204*** (0.073)
Above Median Baseline Math Score	0.186*** (0.071)	0.172** (0.072)	0.185*** (0.069)
Below Median Baseline Math Score	0.171** (0.072)	0.128** (0.064)	0.175** (0.071)
Black, Male	0.163*** (0.060)	0.141*** (0.050)	0.161*** (0.059)
Black, Female	0.175** (0.069)	0.150** (0.069)	0.177** (0.069)
Black, Above Median	0.159** (0.066)	0.145** (0.065)	0.155** (0.065)
Black, Below Median	0.149** (0.069)	0.109* (0.061)	0.154** (0.069)

Notes: Table provides first-stage estimates for Table 1.6. See notes of Table 1.6 for regression details.

Table C.2: Observation Counts for Table 1.6

	Achievement Index (1)	Postsecondary Index (2)	Total Index (3)
Panel A: <i>2SLS Estimates for Pooled Sample</i>			
Observations	12,392	7,715	12,635
Panel B: <i>2SLS Estimates by Subgroup</i>			
Non-Black	1,946	1,037	1,970
Black	10,446	6,678	10,665
Male	5,473	3,200	5,576
Female	6,919	4,515	7,059
Above Median Baseline Math Score	5,790	3,949	5,904
Below Median Baseline Math Score	4,762	2,652	4,877
Black, Male	4,607	2,752	4,690
Black, Female	5,839	3,926	5,975
Black, Above Median	4,484	3,192	4,581
Black, Below Median	4,386	2,520	4,495

Notes: Table provides observation counts for Table 1.6. See notes of Table 1.6 for regression details.

Table C.3: Weak Instruments F Statistics for Table 1.6

	Achievement Index (1)	Postsecondary Index (2)	Total Index (3)
Panel A: <i>2SLS Estimates for Pooled Sample</i>			
F Statistics	8.116	6.543	8.286
Panel B: <i>2SLS Estimates by Subgroup</i>			
Non-Black	12.769	9.070	13.733
Black	6.868	5.858	7.006
Male	8.535	8.892	8.702
Female	7.591	5.309	7.718
Above Median Baseline Math Score	6.976	5.681	7.118
Below Median Baseline Math Score	5.698	3.966	6.055
Black, Male	7.287	7.902	7.402
Black, Female	6.474	4.719	6.567
Black, Above Median	5.701	4.969	5.668
Black, Below Median	4.629	3.154	5.000

Notes: Table reports weak instrument tests using the Kleibergen-Paap Wald F statistic for Table 1.6. See notes of Table 1.6 for regression details.

Table C.4: Outcome Means for Table 1.6

	Achievement Index (1)	Postsecondary Index (2)	Total Index (3)
Panel A: <i>2SLS Estimates for Pooled Sample</i>			
Outcome Mean	0.225	0.014	0.108
Panel B: <i>2SLS Estimates by Subgroup</i>			
Non-Black	0.672	-0.027	0.408
Black	0.163	0.019	0.067
Male	0.152	-0.047	0.044
Female	0.282	0.056	0.158
Above Median Baseline Math Score	0.549	0.071	0.338
Below Median Baseline Math Score	-0.184	-0.075	-0.192
Black, Male	0.084	-0.045	-0.003
Black, Female	0.223	0.063	0.120
Black, Above Median	0.483	0.077	0.295
Black, Below Median	-0.195	-0.068	-0.199

Notes: Table provides outcome means among students not offered a magnet lottery seat for Table 1.6. See notes of Table 1.6 for regression details.

Table C.5: First-Stage Estimates for Table 1.8

	Achievement Testing			
	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: 2SLS Estimates for Pooled Sample				
DPB' ⁰² * 1(Post'02) * 1(Offer)	2.098*** (0.550)	2.105*** (0.552)	1.955*** (0.565)	2.073*** (0.599)
Panel B: 2SLS Estimates by Subgroup				
Black	2.018*** (0.593)	1.999*** (0.591)	1.901*** (0.591)	1.933*** (0.618)
Male	2.653*** (0.550)	2.616*** (0.547)	2.205*** (0.618)	2.265*** (0.619)
Female	1.630*** (0.570)	1.673*** (0.572)	1.723*** (0.551)	1.890*** (0.627)
Above Median Baseline Math Score	2.170*** (0.577)	2.201*** (0.578)	1.977*** (0.596)	2.184*** (0.669)
Below Median Baseline Math Score	1.703** (0.692)	1.704** (0.684)	1.606** (0.796)	1.533* (0.895)
Black, Male	2.626*** (0.574)	2.548*** (0.556)	2.160*** (0.601)	2.185*** (0.608)
Black, Female	1.525** (0.650)	1.551** (0.651)	1.645*** (0.626)	1.706** (0.675)
Black, Above Median	2.139*** (0.652)	2.135*** (0.656)	1.964*** (0.632)	2.111*** (0.715)
Black, Below Median	1.625** (0.732)	1.604** (0.725)	1.492* (0.805)	1.314 (0.904)

Notes: Table provides first-stage estimates for Table 1.8. See notes of Table 1.8 for regression details.

Table C.6: Observation Counts for Table 1.8

	Achievement Testing			
	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>2SLS Estimates for Pooled Sample</i>				
Observations	12,187	12,159	10,565	9,448
Panel B: <i>2SLS Estimates by Subgroup</i>				
Black	10,303	10,302	8,948	7,959
Male	5,359	5,339	4,740	4,128
Female	6,823	6,815	5,821	5,319
Above Median Baseline Math Score	5,712	5,678	4,911	4,306
Below Median Baseline Math Score	4,665	4,676	4,138	3,524
Black, Male	4,515	4,515	3,999	3,476
Black, Female	5,782	5,781	4,944	4,482
Black, Above Median	4,440	4,433	3,824	3,326
Black, Below Median	4,308	4,316	3,810	3,242

Notes: Table provides observation counts for Table 1.8. See notes of Table 1.8 for regression details.

Table C.7: Weak Instruments F Statistics for Table 1.8

	Achievement Testing			
	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>2SLS Estimates for Pooled Sample</i>				
F Statistics	14.534	14.521	11.953	11.967
Panel B: <i>2SLS Estimates by Subgroup</i>				
Black	11.571	11.429	10.358	9.771
Male	23.290	22.912	12.747	13.390
Female	8.187	8.567	9.763	9.100
Above Median Baseline Math Score	14.148	14.499	11.004	10.649
Below Median Baseline Math Score	6.060	6.203	4.068	2.933
Black, Male	20.950	20.973	12.900	12.911
Black, Female	5.498	5.683	6.912	6.384
Black, Above Median	10.758	10.585	9.672	8.713
Black, Below Median	4.928	4.895	3.438	2.113

Notes: Table reports weak instrument tests using the Kleibergen-Paap Wald F statistic for Table 1.8. See notes of Table 1.8 for regression details.

Table C.8: Outcome Means for Table 1.8

	Achievement Testing			
	Reading (1)	Math (2)	Science (3)	Writing (4)
Panel A: <i>2SLS Estimates for Pooled Sample</i>				
Outcome Mean	0.357	0.091	0.071	0.410
Panel B: <i>2SLS Estimates by Subgroup</i>				
Black	0.260	-0.026	-0.038	0.359
Male	0.249	0.045	0.093	0.228
Female	0.443	0.128	0.053	0.550
Above Median Baseline Math Score	0.732	0.554	0.415	0.679
Below Median Baseline Math Score	-0.105	-0.516	-0.357	0.127
Black, Male	0.136	-0.080	-0.035	0.173
Black, Female	0.356	0.016	-0.041	0.501
Black, Above Median	0.628	0.420	0.283	0.638
Black, Below Median	-0.131	-0.539	-0.383	0.119

Notes: Table provides outcome means among students not receiving magnet lottery offer for Table 1.8. See notes of Table 1.8 for regression details.

Table C.9: First-Stage Estimates for Table 1.9

	HS Grad. (1)	College Attendance (18 Months after High School Graduation)			
		Any (2)	2-year (3)	4-year (4)	Top 50 (5)
Panel A: <i>2SLS Estimates for Pooled Sample</i>					
DPB' ⁰² * 1(Post'02) * 1(Offer)	2.211*** (0.544)	2.239*** (0.571)	2.239*** (0.571)	2.239*** (0.571)	2.239*** (0.571)
Panel B: <i>2SLS Estimates by Subgroup</i>					
Black	2.206*** (0.611)	2.138*** (0.641)	2.138*** (0.641)	2.138*** (0.641)	2.138*** (0.641)
Male	2.526*** (0.622)	2.537*** (0.637)	2.537*** (0.637)	2.537*** (0.637)	2.537*** (0.637)
Female	1.790*** (0.548)	1.851*** (0.575)	1.851*** (0.575)	1.851*** (0.575)	1.851*** (0.575)
Above Median Baseline Math Score	2.611*** (0.586)	2.362*** (0.587)	2.362*** (0.587)	2.362*** (0.587)	2.362*** (0.587)
Below Median Baseline Math Score	1.656** (0.726)	1.779** (0.821)	1.779** (0.821)	1.779** (0.821)	1.779** (0.821)
Black, Male	2.600*** (0.674)	2.409*** (0.694)	2.409*** (0.694)	2.409*** (0.694)	2.409*** (0.694)
Black, Female	1.755*** (0.651)	1.801*** (0.650)	1.801*** (0.650)	1.801*** (0.650)	1.801*** (0.650)
Black, Above Median	2.601*** (0.670)	2.277*** (0.675)	2.277*** (0.675)	2.277*** (0.675)	2.277*** (0.675)
Black, Below Median	1.637** (0.748)	1.575* (0.851)	1.575* (0.851)	1.575* (0.851)	1.575* (0.851)

Notes: Table provides first-stage estimates for Table 1.9. See notes of Table 1.9 for regression details.

Table C.10: Observation Counts for Table 1.9

	HS Grad. (1)	College Attendance (18 Months after High School Graduation)			
		Any (2)	2-year (3)	4-year (4)	Top 50 (5)
Panel A: <i>2SLS Estimates for Pooled Sample</i>					
Observations	9,023	7,696	7,696	7,696	7,696
Panel B: <i>2SLS Estimates by Subgroup</i>					
Black	7,809	6,663	6,663	6,663	6,663
Male	3,858	3,191	3,191	3,191	3,191
Female	5,162	4,502	4,502	4,502	4,502
Above Median Baseline Math Score	4,417	3,942	3,942	3,942	3,942
Below Median Baseline Math Score	3,286	2,641	2,641	2,641	2,641
Black, Male	3,333	2,744	2,744	2,744	2,744
Black, Female	4,473	3,916	3,916	3,916	3,916
Black, Above Median	3,565	3,188	3,188	3,188	3,188
Black, Below Median	3,102	2,511	2,511	2,511	2,511

Notes: Table provides observation counts for Table 1.9. See notes of Table 1.9 for regression details.

Table C.11: Weak Instruments F Statistics for Table 1.9

	HS Grad. (1)	College Attendance (18 Months after High School Graduation)			
		Any (2)	2-year (3)	4-year (4)	Top 50 (5)
Panel A: <i>2SLS Estimates for Pooled Sample</i>					
F Statistics	16.529	15.399	15.399	15.399	15.399
Panel B: <i>2SLS Estimates by Subgroup</i>					
Black	13.053	11.120	11.120	11.120	11.120
Male	16.476	15.874	15.874	15.874	15.874
Female	10.677	10.346	10.346	10.346	10.346
Above Median Baseline Math Score	19.872	16.199	16.199	16.199	16.199
Below Median Baseline Math Score	5.196	4.688	4.688	4.688	4.688
Black, Male	14.878	12.033	12.033	12.033	12.033
Black, Female	7.276	7.664	7.664	7.664	7.664
Black, Above Median	15.075	11.395	11.395	11.395	11.395
Black, Below Median	4.792	3.426	3.426	3.426	3.426

Notes: Table reports weak instrument tests using the Kleibergen-Paap Wald F statistic for Table 1.9. See notes of Table 1.9 for regression details.

Table C.12: Outcome Means for Table 1.9

	HS Grad. (1)	College Attendance (18 Months after High School Graduation)			
		Any (2)	2-year (3)	4-year (4)	Top 50 (5)
Panel A: <i>2SLS Estimates for Pooled Sample</i>					
Outcome Mean	0.639	0.652	0.274	0.462	0.005
Panel B: <i>2SLS Estimates by Subgroup</i>					
Black	0.636	0.653	0.278	0.463	0.004
Male	0.591	0.609	0.262	0.423	0.006
Female	0.676	0.683	0.283	0.490	0.004
Above Median Baseline Math Score	0.787	0.710	0.254	0.543	0.007
Below Median Baseline Math Score	0.632	0.570	0.296	0.351	0.001
Black, Male	0.582	0.611	0.266	0.426	0.004
Black, Female	0.676	0.683	0.287	0.489	0.003
Black, Above Median	0.786	0.714	0.258	0.552	0.006
Black, Below Median	0.639	0.575	0.298	0.355	0.001

Notes: Table provides outcome means among students not receiving magnet lottery offer for Table 1.9. See notes of Table 1.9 for regression details.

Chapter 5

Charter Competition Appendix

A Data Appendix

In this section, I cover details of the subjective data cleaning considerations underling my analyses. Table A.1 displays a summary of the source, important variables, year available and any miscellaneous notes for the data used in this paper.

A.1 ODE Restricted-Access Staff Data

In the staffing data, a teacher’s education category can only take the following values: Non-Degree, Associate, BA, MA, Education Specialist, PhD, Other, less than HS Diploma, HS Diploma, and GED.

I only keep staff classified as “regular teachers” receiving annual salaries, with previous education categories: non-degree, BA, and MA. I truncate the sample and drop teacher-year observations with real (\$2010) annual salaries less than \$15,000 or greater than \$85,000 as well as teachers reporting a null value for full-time equivalency units.

The next issue with these data is that teacher experience doesn’t always increment properly. To fix this, I use the experience from the first year a teacher is observed in the dataset and then increment for all subsequent years that the teacher is observed in the data. For some teachers, they have valid experience for years later in their career. This information is utilized by decrementing to fill previous missing years. In the case where the incremented and decremented imputed experience disagree, I take the values of the decremented experience because they appear more accurate. Also, because 2012 is missing, but 2013 and 2014 are available, for teachers observed teaching after 2012, experience is incremented assuming they also taught in 2012.

A.2 SERB Collectively Bargained Contract Data

First, I only keep contracts from full-time, regular teachers. In Ohio, the school year typically runs from the end of August to the beginning of June, so for mid-year negotiations, I consider all contracts that begin being enforced from January through May as belonging to the current school year. For example, a contract with an enforcement date in February 2010 would be considered to be effective during the 2009-10 school year. Outliers in nominal entry- and top-level salaries are dropped. Specifically, I code as missing contract-education observations with starting salaries less than \$10,000 and top salaries greater than \$200,000. I also recode as missing the number of salary steps to reach the top of a pay scale for contracts with 40 or more steps. Finally, if the total number of years required to ascend the pay scale was greater than 50, I set the variable capturing the number of years between each step to missing.¹ It is also worth noting that all monetary variables are brought into real 2010\$ by using the BEA Personal consumption expenditures from Table 1.1.4. “Price Indexes for Gross Domestic Product”.

A.3 Other Sample Restrictions

For my final analysis sample, I exclude any districts that are entirely made up of charter schools. In Ohio, charters are recorded as being in their own local area agency so this effectively limits the analysis to all non-chartered districts. I also drop special needs and other non-traditional regional education service agencies (including joint vocational districts) so that the analysis is only performed on traditional K-12 school districts. Lastly, the sample is limited to districts that report having more than 50 students enrolled and more than 5 teachers employed within the district.

¹The total number of years required to ascend the pay scale is defined as the number of pay scale steps \times number of years between each pay scale step.

Table A.1: Data Summary

Provider	Source	Relevant Variables	Years	Miscellaneous
Ohio Department of Education	District Foundation Settlement Reports: Community School Deduction	# Students Transferred to B&M and Digital Charters	2001-2012	<ul style="list-style-type: none"> N/A
Ohio Department of Education	Restricted Access Staff Data	Teacher Education; Teacher Experience; School/District Employer (Including Charters); Payments; Teacher Transfers	1996-2011, 2013-2014	<ul style="list-style-type: none"> Salary includes additional compensation e.g., coaching positions. Experience doesn't increment properly, see Appendix A.1.
Ohio Department of Education	ODE-Advanced Reports: District Enrollment	Total district enrollment	1995-2011	<ul style="list-style-type: none"> N/A
Ohio Department of Education	ODE-Advanced Reports: School Ratings	% AYP indicators met; rating (e.g., "Academic Watch")	1998-2011	<ul style="list-style-type: none"> N/A
Ohio State Employment Relations Board	SERB Contract Clearinghouse	Entry/Top-level Salaries by education category; Pay scale steps; Contract start dates	1982-2012	<ul style="list-style-type: none"> Top-level salaries are actually the first year that an additional year of experience does not increase the salary earned, regardless if pay would ever increase with more experience.
Ohio Department of Taxation	School District Property Tax Database	Total and residential district-level property values	1986-2013	<ul style="list-style-type: none"> N/A
National Center of Educational Statistic's Common Core of Data	School District Universe Survey	Special Education Enrollment; Number of Teachers Employed	1987-2011	<ul style="list-style-type: none"> N/A
National Center of Educational Statistic's Common Core of Data	School Building Universe Survey	Black student enrollment; Free/Reduced price lunch eligible student enrollment	1987-2011	<ul style="list-style-type: none"> In 2007, free-lunch eligible student enrollment is unavailable.
National Center of Educational Statistic's Common Core of Data	School District Finance Survey	Aggregated and disaggregated revenues and expenditures; Payments to charter schools	1989-2011	<ul style="list-style-type: none"> N/A

Table A.2: Example of Teacher Salary Contract

Experience	Non-Degree	BA	BA+150	MA	MA+30
0	22,013	25,898	26,934	28,488	29,524
1	22,790	26,934	28,099	29,783	30,948
2	24,085	27,970	29,265	31,078	32,373
⋮	⋮	⋮	⋮	⋮	⋮
7	28,229	33,149	35,092	37,552	39,494
8	29,265	34,185	36,257	38,847	40,919
9	29,265	35,221	37,423	40,142	42,343
10	29,265	36,257	38,588	41,437	43,768
⋮	⋮	⋮	⋮	⋮	⋮
13	29,265	39,365	42,084	45,322	48,041
14	30,042	40,401	43,250	46,616	49,485
15	30,042	40,401	43,250	46,616	49,485
⋮	⋮	⋮	⋮	⋮	⋮
28	30,042	45,348	49,014	52,067	54,578

Notes: This table presents a fictitious collectively bargained teacher's salary matrix. A teacher's pay can be determined simply by referring to the number of years they have taught in the district and the level of education they possess. Education categories are respectively no degree, Bachelor's degree, Bachelor's degree with 150 additional credit hours, Master's degree, and Master's degree with 30 additional credit hours. SERB data contain starting-level salary information for all education levels (i.e., all entries corresponding to the null experience level). While data sometimes exist for top-level salaries (i.e., all entries for 28 years of experience), SERB data custodians often instead code top-level salaries as the first experience step in which subsequent experience gains incurs no additional salary premium. The bold values represent these incorrectly-coded top-level salaries for the example contract.

B Full Salary Distribution Imputation

By making use of both my teacher-level micro data as well as information about contract negotiation years, I am able to approximate negotiated salaries for each step of the pay scale. This allows me to estimate the effect of charter competition along the entire negotiated pay scale distribution. Most importantly, these teacher-level data include the college degree, number of years in the district, and annual compensation. This is enough to identify which step on the pay scale that each teacher should in theory be compensated.

Next, I collapse the teacher data so that the unit of observation is the district, year, college degree, experience level and calculate the median payments within each cell.² I then merge in contract negotiation dates from my SERB data so that observations are limited to the years in which a new contract is being enforced. In the end I want a dataset that has approximated salary steps for each district, year, education, and experience combination, but teachers are not observed teaching in each of these combinations. To solve this problem, I impute district-specific annual payments for any missing pay scale steps by using predicted values from the regression

$$\begin{aligned} \text{Median Real Payment}_{idet} = & \alpha + \gamma_{it} * \text{Experience}_{idet} + \delta_{it} * \text{Experience}_{idet}^2 \\ & + \phi_{it} + \theta_d + \epsilon_{iset} , \end{aligned} \quad (\text{B.1})$$

where Median Real Payment_{idet} is the 2010-inflation-adjusted median annual payment to teachers in district i , with degree d , years of experience e , during the school year in which the contract for the district is enforced t . I also include a district-year-specific quadratic for a continuous measure of experience (Experience_{idet}) and include the main effects ϕ_{it} . θ_d are degree fixed effects.

Equation (B.1) is simply fitting a district-year-specific quadratic experience term (the γ_{it} 's and δ_{it} 's) to the median payments within each district, during the contract's enforcement year, and allows for level shifts in the pay scale for each education category.

I then create a balanced panel of all possible district-year-education-experience cells and for cells without actual payment information I impute using predicted values from (B.1). This dataset is used for the analysis explained in Section 2.6 that generates Figure 2.5.

C Alternative Measures of Charter Competition

In this section, I provide estimates for the main revenue and expenditure outcomes of this study using four alternate measures of charter competition. The first two measures are respectively the number of charter buildings in operation during the given school year that are within 5 and 10

²This helps exclude outlier salaries due to some teachers receiving bonus payments for additional responsibilities such as coaching an athletic team.

miles of the TPSD's geographic center.³ It is likely that parents in more rural areas of the state would be willing to have their child travel further distances to attend a school. As a result, I also create a measure of charter competition based on the overall district size. Specifically, for each TPSD, I have shape files containing coordinates that trace out each enrollment boundary. I identify 20 equidistant coordinates along with the coordinates corresponding to points on the opposite side of the district. I calculate the distance between these 20 points and their accompanying opposite points and take half of this average to calculate my measure of average TPSD radius. Finally, I count the number of charter schools open within 150 percent of this metric. The final measure of charter competition is simply the number of operating charters that are located within district boundaries.

I present the robustness of my main results across these measures of charter competition in Table C.1. For comparison, Column 1 provides estimates using my main competition metric. Overall, my results are extremely robust across competition measures showing that the effects I find are not simply an artifact of my particular competition metric. Estimates measuring competition as the number of operating chartering within 5 miles of the TPSD district center (column 2) most closely mirror my main results. As the radius is extended (columns 3 and 4) estimates attenuate slightly. One possible explanation is that as the boundaries increase, charters are counted that may not actually be viable options for parents in the district. Measuring charter competition as the number of charters operating within TPSD boundaries (column 5) attenuates estimates the most. This could happen if charter competition outside of TPSD boundaries present particularly strong competitive pressures. However, even using this measure, all of the effect signs remain the same making the overall story unchanged. Finally, it is worth noting that each of these alternative measures miss the digital charter competition that is captured by my preferred measure.

³TPSD center coordinates are provided in the NCES Common Core of Data.

Table C.1: Effect of Various Charter Competition Measures on Key IV Estimates

	Original Measure	Alternative Measures			
	Fraction of Students attending Charter \times 100 (1)	# Charters within 5 miles of TPSD center (2)	# Charters within 10 miles of TPSD center (3)	# Charters within $1.5 \times$ TPSD average radius (4)	# Charters within TPSD boundary (5)
Panel A: TPSD Revenues					
IHS of Total	-0.018*** (0.005)	-0.018*** (0.006)	-0.011*** (0.004)	-0.011*** (0.004)	-0.006** (0.003)
IHS of Federal	-0.041*** (0.007)	-0.039*** (0.009)	-0.022*** (0.007)	-0.021*** (0.006)	-0.011** (0.005)
IHS of Local	-0.034*** (0.006)	-0.034*** (0.008)	-0.020*** (0.005)	-0.019*** (0.006)	-0.010*** (0.004)
IHS of Residential Values	-0.026*** (0.005)	-0.028*** (0.006)	-0.016*** (0.004)	-0.017*** (0.005)	-0.010*** (0.004)
Panel B: Union Salary					
Log of Entry	-0.002 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.000 (0.002)	-0.000 (0.001)
Log of Imputed Top	-0.010** (0.004)	-0.007 (0.005)	-0.005 (0.003)	-0.004 (0.003)	-0.002 (0.001)
Panel C: TPSD Expenditures					
IHS of Total (Excl. Charter Transfers)	-0.017*** (0.006)	-0.020 (0.012)	-0.013 (0.008)	-0.014 (0.009)	-0.005 (0.003)
IHS of Instructional	-0.023*** (0.004)	-0.013** (0.006)	-0.008** (0.004)	-0.009** (0.005)	-0.003 (0.002)
IHS of Capital Outlays	0.073** (0.034)	0.103 (0.080)	0.066 (0.052)	0.069 (0.056)	0.022 (0.022)
IHS of New Construction Capital Outlays	0.113 (0.114)	0.142 (0.272)	0.090 (0.181)	0.091 (0.195)	0.027 (0.079)

Notes: Standard Errors in parentheses are clustered by district. See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). Refer to Tables 2.4, 2.5, and 2.6 for specification details for panels A, B, and C respectively. This table reports 2SLS estimates of the effect of various measures of charter competition on key outcomes. In Column 1, competition is measured as the fraction of the district's membership attending charter schools times 100. Columns 2 and 3 present estimates of competition as measured by the number of charters open within 5 and 10 miles of the TPSD geographic center. In Column 4, competition is measured as the number of charters open within 1.5 times the district's average radius. In Column 5, competition is measured as the number of charters open within the TPSD's boundaries.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

D Ohio School Rating Designation System

Ohio uses four measures to evaluate the performance of schools and districts: (1) State Indicators, (2) Performance Index, (3) Adequate Yearly Progress (AYP), and (4) Value-Added data.

D.1 State Indicators

Schools and districts are evaluated based on 26 measures against state determined goals, as seen in Table D.1.

D.2 Performance Index Scores

The Performance Index score is a continuous measure that summarizes student achievement (for all students, not just those testing proficient or higher) at the school or district level into a single index. These scores are calculated using a weighted average of individual student performance levels on proficiency and achievement tests. The weights for different score levels can be found in Table D.2. You can see that there are large penalties to testing below proficiency, but only small bonuses for testing above the “Proficient level”. Thus, it takes several high scores to balance a single low score.

Each weighted score is multiplied by the percentage of students scoring at that level. In order to generate these percentages, schools count the number of students present during the October count and 120 consecutive school days of enrollment, including the March testing period. Other rules apply to counting students that have a small impact on student enrollment counts.⁴ Specifically, the Performance Index rating is calculated for school/district i in academic year t using the previous year’s $(t - 1)$ academic scores as follows:

$$PI_{it} = 1.2 (\% \text{ Advanced}_{i,t-1}) + 1.1 (\% \text{ Accelerated}_{i,t-1}) + 1.0 (\% \text{ Proficient}_{i,t-1}) + 0.6 (\% \text{ Basic}_{it}) + 0.3 (\% \text{ Limited}_{i,t-1}). \quad (\text{D.1})$$

D.3 School and District Rating Calculations

Once Performance index scores are calculated, school and districts are then categorized into five levels as seen in Table D.3.

In addition to these raw PI scores, schools and districts can also experience a boost in their rating based on certain growth calculations. Specifically a school/district can move from “Emergency”

⁴e.g., 1 percent limitation on counting alternately assessed students as proficient. The total count of all student scores includes up to five tests per student.

to “Watch” or from “Watch” to “Improvement” if: (1) the PI score improved in each of the last two years *and* (2) the total two-year gain was at least 10 points *and* (3) a gain of at least 3 points took place in the current year.

D.4 Value Added

Value-Added (VA) measures try to account for how much progress and growth has been made since the previous year. These VA measures are calculated for grades 4-8 in reading and math. A composite is made of these scores and schools are delineated as performing above, at, or below expectations. Schools performing above expectation may increase its designation by one rating.

D.5 Ratings Determined by Number of AYP Indicators met

The building/district designation will always be the higher category determined by the PI score or the number of AYP indicators met. Ohio schools that meet AYP must be designated as Continuous Improvement or higher. Conversely, schools not meeting AYP for three years in a row and not meeting it for more than one student group in the most recent year can be rated no higher than Continuous Improvement. AYP indicators are based on 3-8th grade reading and math assessments as well as 10th grade Ohio graduation testing in reading and math.

Schools meet AYP when: (1) The proficiency level weighted across all tested grades is at or above the AYP goal, (2) If the above proficiency level is met when combined with the previous year, (3) a student group must make a 10 percent or greater reduction in its percentage of non-proficient students from the previous year, and they must meet the AYP goal in the secondary indicator (graduation rate and/or attendance rate), or (4) through the growth model, i.e., a non-proficient student projected to be on a path to proficiency within two years will be treated as proficient in the current year.⁵ The growth model uses data from the Ohio Achievement tests in grades 3-8 so traditional high school buildings (those with grades 9-12) cannot use the growth model to meet AYP (ODE, 2011). Figure D.1 provides a graphical summary of this information.

⁵Student groups include: All students, Black, American Indian, Asian, Hispanic, Multi-Racial, White, Economically Disadvantaged, Limited English Proficiency, and Students with Disabilities.

Table D.1: State Indicators and 2010-11 Goals

State Indicator	'10-'11 Goal	# of indicators
3rd-grade achievement tests: reading and math	75%	2
4th-grade achievement tests: reading and math	75%	2
5th-grade achievement tests: reading, math, and science	75%	3
6th-grade achievement tests: reading and math	75%	2
7th-grade achievement tests: reading and math	75%	2
8th-grade achievement tests: reading, math, and science	75%	3
Ohio Graduation Test- 10th-grade: reading, math, writing, science, social studies	75%	5
Ohio Graduation Test- 11th-grade: reading, math, writing, science, social studies	85%	5
Graduation Rate	90%	1
Attendance Rate	93%	1
Total		26

Source: Ohio Department of Education Report Card Guide 2010-11 – ([Link](#))

Table D.2: Performance Index Scores and Report Card Designation

Performance Level	Weight
Untested Student	0
Below Basic	0.3
Basic	0.6
Proficient	1.0
Accelerated	1.1
Advanced	1.2

Source: Ohio Department of Education
Performance Index Calculator – ([Link](#))

Table D.3: Performance Index Scores and Report Card Designation

PI Score	Report Card Designation
0-69	Academic Emergency
70-79	Academic Watch
80-89	Continuous Improvement
90-99	Effective
100-120	Excellent

Source: Ohio Department of Education
Performance Index Calculator – [\(Link\)](#)

Figure D.1: Summary of Rating Designation

Indicators Met		Performance Index Score		AYP Status	Preliminary Designation	Did the Preliminary Designation increase or decrease based on the AYP Status?	IF YES <u>STOP HERE</u> No additional change to the designation can occur based on the value added calculation	IF NO <u>CONTINUE</u> Value-added <u>MAY</u> affect a designation when it has not been changed by the AYP Status	Preliminary Designation		Amount of Growth Using Value-Added Calculation	Final Designation	
94% - 100%	or	100 to 120	and	Met or Not Met	Excellent				Excellent	and	Above expected growth	Excellent with Distinction	
											Below expected growth for at least 3 consecutive years	Effective	
											Otherwise no effect on rating	Excellent	
75% - 93.9%	or	90 to 99.9	and	Met or Not Met	Effective				Effective	and	Above expected growth	Excellent	
											Below expected growth for at least 3 consecutive years	Continuous Improvement	
											Otherwise no effect on rating	Effective	
0% - 74.9%	or	0 to 89.9	and	Met	Continuous Improvement				Continuous Improvement	and	Above expected growth	Effective	
												Below expected growth for at least 3 consecutive years	Academic Watch
												Otherwise no effect on rating	Continuous Improvement
50% - 74.9%	or	80 to 89.9	and	Not Met		Academic Watch	and	Above expected growth	Continuous Improvement				
								Below expected growth for at least 3 consecutive years	Academic Emergency				
								Otherwise no effect on rating	Academic Watch				
31% - 49.9%	or	70 to 79.9	and	Not Met	Academic Watch	Academic Emergency	and	Above expected growth	Academic Watch				
								Otherwise no effect on rating	Academic Emergency				
0% - 30.9%	and	0 to 69.9	and	Not Met	Academic Emergency								

Source: Ohio Department of Education Report Card Guide 2010-11 (ODE, 2011) – [\(Link\)](#)

E Robustness: No Child Left Behind & Great Recession

First, I attempt to account for NCLB contamination by adding additional controls to equation (2.2). Specifically, I add indicator variables for whether the district has had an AYP failure spell for 1-2 and 3+ consecutive years, respectively. Because AYP measures were created by the ODE even for years prior to the NCLB introduction, I also interact these binaries with the same $t - 2$ post-2002 school year binaries as I do for the main instruments. This attempts to control for the differential change in response to consecutive AYP failure that occurs at 2002, precisely the year of the introduction of one of the policies determining charter entry. Tables E.1 to E.2 display the main sets of results for this specification. I find that results are almost identical to the specifications omitting these NCLB controls.

Second, many of the NCLB sanctions only could take effect once a school/district failed AYP for two consecutive years. As a result, outcomes from 2004 and earlier are plausibly unaffected by NCLB policies. Tables E.3 to E.4 present estimates of the main regressions from equations (2.1) and (2.2), but limit the regression samples to school years 2004 and earlier. Because the Great Recession occurred after this cutoff, these regressions also test whether the recession drives my findings. Again, while the estimates are often less precise due to the loss of seven years of data, the main story still remains in tact. The robustness of my main estimates to these tests suggests that I am able to adequately control for any outside influence of NCLB on district resource acquisition and allocation.

Table E.1: Effect of Charter Transfers on District Revenues (AYP Interaction Controls)

Panel A: IHS of Total Revenues	Total	Federal	Local
	(1)	(2)	(3)
Fraction Charter Transfers $\times 100$ – OLS	-0.006*** (0.002)	-0.025*** (0.005)	-0.018*** (0.003)
Fraction Charter Transfers $\times 100$ – IV	-0.017*** (0.006)	-0.036*** (0.008)	-0.029*** (0.006)
Panel B: IHS of Federal Revenues	Child Nutrition Act	Disabilities Act	Other
	(5)	(6)	(7)
Fraction Charter Transfers $\times 100$ – OLS	-0.035*** (0.012)	-0.058*** (0.021)	-0.009* (0.005)
Fraction Charter Transfers $\times 100$ – IV	-0.050** (0.020)	-0.053 (0.084)	-0.001 (0.011)
Panel C: IHS of Local Revenues	Property Tax	School Lunch	Other
	(8)	(9)	(10)
Fraction Charter Transfers $\times 100$ – OLS	-0.018*** (0.003)	-0.056*** (0.008)	-0.006 (0.007)
Fraction Charter Transfers $\times 100$ – IV	-0.024*** (0.005)	-0.052*** (0.017)	-0.035** (0.017)
Panel D: Property Tax Decomposition	IHS of Property Value		
	Total	Residential	Millage
	(11)	(12)	(13)
Fraction Charter Transfers $\times 100$ – OLS	-0.022*** (0.003)	-0.016*** (0.003)	0.014 (0.085)
Fraction Charter Transfers $\times 100$ – IV	-0.023*** (0.005)	-0.024*** (0.005)	-0.194 (0.132)

Notes: N= 11,449 district-year observations. Standard Errors in parentheses are clustered by district. First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{t-2} * 1(\text{Acad. E.})_{t-2} = 2.158^{***}$ (0.659); Post 2002 $_{t-2} * 1(\text{Acad. W.})_{t-2} = 3.124^{***}$ (0.624); and $t - 1$ Char. Elig. (Urban 8/21) = 2.468*** (0.906). See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). The F statistic for excluded instruments is 8.909***. This table reports OLS (see equation (2.1)) and 2SLS (see equation (2.2)) estimates for the effect of charter competition on district revenues. The endogenous variable is the fraction of the district's membership attending charter schools times 100. Each regression also includes two binaries one for whether during school-year $t - 2$, the district had missed AYP for 1-2 and 3+ consecutive years, respectively. These binaries are also interacted with an indicator equal to one if the given $t - 2$ school-year was after 2002. Each cell provides the result of a separate regression. See Table H.2 for the mean of each dependent variable and Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table E.2: Effect of Charter Transfers on District Expenditures (AYP Interaction Controls)

	IHS of Expenditure			IHS of Capital Outlays		
	Charter Payments (100,000s)	Total (Net of Charter Payment)	Instruction	Capital Outlays	Other	New Construction
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Charter Transfers $\times 100 - \text{OLS}$	1.148*** (0.260)	-0.007*** (0.002)	-0.020*** (0.004)	0.069*** (0.016)	-0.020*** (0.003)	0.137*** (0.048)
Fraction Charter Transfers $\times 100 - \text{IV}$	2.180*** (0.461)	-0.016*** (0.007)	-0.020*** (0.004)	0.071* (0.037)	-0.027*** (0.004)	0.100 (0.124)
						0.001 (0.021)

Notes: N= 11,449 district-year observations. Standard Errors in parentheses are clustered by district. See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{t-2} * \mathbb{1}(\text{Acad. E.})_{t-2} = 2.158^{***}$ (0.659); Post 2002 $_{t-2} * \mathbb{1}(\text{Acad. W.})_{t-2} = 3.124^{***}$ (0.624); and $t - 1$ Char. Elig. (Urban 8/21) = 2.468*** (0.906). The F statistic for excluded instruments is 8.909***. This table reports OLS and 2SLS estimates of the effect of charter competition on different forms of teacher mobility. The endogenous variable is the fraction of the district's membership attending charter schools times 100. Each regression includes district and commute-zone-by-contract-start-year fixed effects. Each Panel and column provide the results of a separate regression. Each regression also includes two binaries one for whether during school-year $t - 2$, the district had missed AYP for 1-2 and 3+ consecutive years, respectively. These binaries are also interacted with an indicator equal to one if the given $t - 2$ school-year was after 2002. See Table H.2 for the mean of each dependent variable and Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table E.3: Effect of Charter Transfers on District Revenues (Pre-2005 Sample)

Panel A: IHS of Total Revenues	Total	Federal	Local
	(1)	(2)	(3)
Fraction Charter Transfers $\times 100$ – OLS	-0.002 (0.002)	-0.013*** (0.005)	-0.009*** (0.003)
Fraction Charter Transfers $\times 100$ – IV	-0.022** (0.009)	-0.038*** (0.012)	-0.032*** (0.010)
Panel B: IHS of Federal Revenues	Child Nutrition Act	Disabilities Act	Other
	(5)	(6)	(7)
Fraction Charter Transfers $\times 100$ – OLS	-0.009 (0.009)	-0.024 (0.018)	-0.001 (0.004)
Fraction Charter Transfers $\times 100$ – IV	-0.030 (0.023)	-0.053 (0.102)	-0.005 (0.012)
Panel C: IHS of Local Revenues	Property Tax	School Lunch	Other
	(8)	(9)	(10)
Fraction Charter Transfers $\times 100$ – OLS	-0.010*** (0.003)	-0.036*** (0.005)	0.002 (0.008)
Fraction Charter Transfers $\times 100$ – IV	-0.021*** (0.007)	-0.042** (0.018)	-0.067** (0.026)
Panel D: Property Tax Decomposition	IHS of Property Value		
	Total	Residential	Millage
	(11)	(12)	(13)
Fraction Charter Transfers $\times 100$ – OLS	-0.024*** (0.004)	-0.011*** (0.004)	0.098 (0.163)
Fraction Charter Transfers $\times 100$ – IV	-0.027*** (0.006)	-0.028*** (0.007)	-0.158 (0.198)

Notes: N= 7,213 district-year observations. Standard Errors in parentheses are clustered by district. First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{\tau-2} * 1(\text{Acad. E.})_{\tau-2} = 2.475^{***}$ (0.543); Post 2002 $_{\tau-2} * 1(\text{Acad. W.})_{\tau-2} = 2.207^{**}$ (1.125); and $t - 1$ Char. Elig. (Urban 8/21) = 1.626*** (0.588). See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). The F statistic for excluded instruments is 9.551***. This table reports OLS (see equation (2.1)) and 2SLS (see equation (2.2)) estimates for the effect of charter competition on district revenues. The endogenous variable is the fraction of the district's membership attending charter schools times 100. The sample is additionally restricted to school-years prior to 2005, the first year when AYP discipline could be enforced. Each cell provides the result of a separate regression. See Table H.2 for the mean of each dependent variable and Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table E.4: Effect of Charter Transfers on District Expenditures (Pre-2005 Sample)

	IHS of Expenditure			IHS of Capital Outlays		
	Charter Payments (100,000s)	Total (Net of Charter Payment)	Instruction	Capital Outlays	Other	New Construction
	(1)	(2)	(3)	(4)	(5)	(6)
						(7)
Fraction Charter Transfers $\times 100 - \text{OLS}$	0.715* (0.407)	-0.004 (0.003)	-0.008** (0.003)	0.058*** (0.019)	-0.019** (0.008)	0.073* (0.039)
Fraction Charter Transfers $\times 100 - \text{IV}$	1.431*** (0.469)	-0.014 (0.010)	-0.010** (0.005)	0.084 (0.059)	-0.032*** (0.008)	0.144 (0.198)
						0.017 (0.028)

Notes: N= 7,213 district-year observations. Standard Errors in parentheses are clustered by district. See footnote 28 on page 68 for details on the inverse hyperbolic sine transformation (IHS). First-stage estimates (and standard errors) for excluded instruments are: Post 1999 $_{t-2} * \mathbb{1}(\text{Acad. E.})_{t-2} = 2.475^{***}$ (0.543); Post 2002 $_{t-2} * \mathbb{1}(\text{Acad. W.})_{t-2} = 2.207^{***}$ (1.125); and $t - 1$ Char. Elig. (Urban 8/21)=1.626*** (0.588). The F statistic for excluded instruments is 9.551***. This table reports OLS and 2SLS estimates of the effect of charter competition on different forms of teacher mobility. The endogenous variable is the fraction of the district's membership attending charter schools times 100. Each regression includes district and commute-zone-by-contract-start-year fixed effects. Each Panel and column provide the results of a separate regression. The sample is additionally restricted to school-years prior to 2005, the first year when AYP discipline could be enforced. See Table H.2 for the mean of each dependent variable and Table H.1 for tests of overidentification. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

F Mechanical Bias for Models with Partially Fixed Dependent Variables

Consider an outcome y that can only vary intermittently. Suppose that if y was able to vary annually, the true model would be given by

$$y_{it} = \beta_0 + \beta_1 x_{it} + \epsilon_{it}, \quad (\text{F.1})$$

but y can only vary intermittently, so instead the econometrician only observes

$$y_{it}^* = \begin{cases} y_{i,t-g} & \text{if } \mathbb{1}(\text{fixed}_{it}) = 1 \\ y_{it} & \text{o.w.} \end{cases}$$

where g is the number of periods since the last negotiation for district i during school year t . $\mathbb{1}(\text{fixed}_{it})$ is an indicator variable equal to one during years in which the y value is fixed for the given district. Thus, you can think of the measurement error term, ν as

$$\nu_{it} = \begin{cases} y_{it} - y_{i,t-g} & \text{if } \mathbb{1}(\text{fixed}_{it}) = 1 \\ 0 & \text{o.w.} \end{cases}$$

We can now rewrite (F.1) as

$$\begin{aligned} y_{it}^* + \nu_{it} &= \beta_0 + \beta_1 x_{it} + \epsilon_{it} \\ y_{it}^* + \mathbb{1}(\text{fixed}_{it})(y_{it} - y_{i,t-g}) &= \beta_0 + \beta_1 x_{it} + \epsilon_{it} \\ y_{it}^* &= \beta_0 + \beta_1 x_{it} + \underbrace{\epsilon_{it} - \mathbb{1}(\text{fixed}_{it})(y_{it} - y_{i,t-g})}_{\equiv \xi_{it}} \end{aligned}$$

But now we can rewrite ξ_{it} by substituting back in (F.1) to get

$$\begin{aligned} \xi_{it} &= \epsilon_{it} - \mathbb{1}(\text{fixed}_{it}) [\beta_0 + \beta_1 x_{it} + \epsilon_{it} - \beta_0 - \beta_1 x_{i,t-g} - \epsilon_{i,t-g}] \\ &= \epsilon_{it} - \mathbb{1}(\text{fixed}_{it}) [\beta_1 \Delta_g x_{it} + \Delta_g \epsilon_{it}] \end{aligned}$$

where $\Delta_g z_{it} \equiv z_{it} - z_{i,t-g}$.

Then assessing consistency we see that

$$\begin{aligned}
plim \hat{\beta}_1 &= \beta_1 + \frac{Cov(x_{it}, \xi_{it})}{Var(x_{it})} \\
&= \beta_1 + \frac{Cov(x_{it}, \epsilon_{it} - \mathbb{1}(fixed_{it}) [\beta_1 \Delta_g x_{it} + \Delta_g \epsilon_{it}])}{\sigma_x^2} \\
&= \beta_1 + \underbrace{\frac{Cov(x_{it}, \epsilon_{it})}{\sigma_x^2}}_{\rightarrow 0} - \frac{Cov(x_{it}, \mathbb{1}(fixed_{it}) [\beta_1 \Delta_g x_{it}])}{\sigma_x^2} \\
&\quad - \underbrace{\frac{Cov(x_{it}, \mathbb{1}(fixed_{it}) [\Delta_g \epsilon_{it}])}{\sigma_x^2}}_{\rightarrow 0}
\end{aligned} \tag{F.2}$$

where I assume that $\mathbb{1}(fixed)_{it}$ is independent of $x_{i,t}$, $x_{i,t-g}$, ϵ_{it} and $\epsilon_{i,t-g}$ so that $\mathbb{1}(fixed)_{it}$ can be factored out. For the final term to go to zero, we must further suppose that $x_{i,t}$ is independent of all lagged errors. Then under these assumptions

$$\frac{Cov(x_{it}, \mathbb{1}(fixed_{it}) [\Delta_g \epsilon_{it}])}{\sigma_x^2} = \frac{Cov(x_{it}, \mathbb{1}(fixed_{it}) \epsilon_{it})}{\sigma_x^2} + \frac{Cov(x_{it}, \mathbb{1}(fixed_{it}) \epsilon_{i,t-g})}{\sigma_x^2} \rightarrow 0$$

Thus, (F.2) can be rewritten as

$$\begin{aligned}
plim \hat{\beta}_1 &= \beta_1 - \frac{Cov(x_{it}, \beta_1 \cdot \mathbb{1}(fixed_{it}) [\Delta_g x_{it}])}{\sigma_x^2} \\
&= \beta_1 \left(1 - \frac{Cov(x_{it}, \cdot \mathbb{1}(fixed_{it}) [\Delta_g x_{it}])}{\sigma_x^2} \right)
\end{aligned}$$

This expression can be simplified. Notice that

$$\begin{aligned}
Cov(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\
Cov(x_{it}, \mathbb{1}(fixed_{it})x_{it}) &= \mathbb{E}(x_{it} \cdot \mathbb{1}(fixed_{it})x_{it}) \\
&\quad - \mathbb{E}(x_{it}) \mathbb{E}(\mathbb{1}(fixed_{it})x_{it}) \\
&= \delta \mathbb{E}(x_{it}^2) - \delta \mathbb{E}(x_{it})^2 = \boxed{\delta \sigma_x^2}
\end{aligned}$$

$$\begin{aligned}
Cov(x_{it}, \mathbb{1}(fixed_{it})x_{i,t-g}) &= \mathbb{E}(x_{it} \cdot \mathbb{1}(fixed_{it})x_{i,t-g}) \\
&- \mathbb{E}(x_{it}) \mathbb{E}(\mathbb{1}(fixed_{it})x_{i,t-g}) \\
&= \delta \mathbb{E}(x_{it} \cdot x_{i,t-g}) - \delta \mathbb{E}(x_{it}) \mathbb{E}(x_{i,t-g}) \\
&= \delta \underbrace{[\mathbb{E}(x_{it} \cdot x_{i,t-g}) - \mathbb{E}(x_{it}) \mathbb{E}(x_{i,t-g})]}_{=Cov(x_{it}, x_{i,t-g}) \equiv \sigma_{\Delta x}} = \boxed{\delta \cdot \sigma_{\Delta x}}
\end{aligned}$$

where I denote $\delta \equiv \mathbb{E}(\mathbb{1}(fixed_{it}))$, i.e., the fraction of observations that are fixed in t .

Now I will show that $\frac{Cov(x_{it}, \mathbb{1}(fixed_{it})[\Delta_g x_{it}])}{\sigma_x^2} \in [0, 2]$.

$$\begin{aligned}
\frac{Cov(x_{it}, \mathbb{1}(fixed_{it})[\Delta_g x_{it}])}{\sigma_x^2} &= \frac{Cov(x_{it}, \mathbb{1}(fixed_{it})x_{it})}{\sigma_x^2} - \frac{Cov(x_{it}, \mathbb{1}(fixed_{it})x_{i,t-g})}{\sigma_x^2} \\
&= \delta \frac{\sigma_x^2}{\sigma_x^2} - \delta \frac{\sigma_{\Delta x}}{\sigma_x^2} \\
&= \delta \left(1 - \underbrace{\frac{\sigma_{\Delta x}}{\sigma_x^2}}_{\rho_x(g)} \right) \\
&= \underbrace{\delta}_{\in [0,1]} \left(1 - \underbrace{\rho_x(g)}_{\in [-1,1]} \right) \in [0, 2]
\end{aligned}$$

where $\rho_x(g)$ is the autocorrelation function for g lags and is obtained by assuming stationarity in x .⁶ Finally, we see that

$$\boxed{plim \hat{\beta} = \beta_1 [1 - \delta (1 - \rho_x(g))]} \quad (\text{F.3})$$

Thus, the sign of the bias is dependent on the sign of the autocorrelation function $\rho_x(g)$.

F.1 Monte Carlo Simulations

In this section, I provide Monte Carlo Evidence of the bias formula in (F.3). To do this, I generate 3 sets of data with varying levels of serial correlation. For the $\rho_x(3) = 1$ case, x is a 10,000 observation array of sequential positive integers. For $\rho_x(3) = 0.85$, x follows an AR(1) process yielding the given serial correlation. For $\rho_x(3) = 0$, x is drawn from a uniform distribution

⁶The autocorrelation function provides the correlation coefficient for a given variable between two different periods of time. Refer to [Okui \(2014\)](#) for estimators of autocovariance and autocorrelations for panel data with individual and time effects.

with values from 0 to 100. For each of these x variables, I allow y only to vary every third observation (i.e., $\delta = 0.66\bar{7}$). For the observations when y can vary, I set y using

$$y_t = 10 + 1 \cdot x_t + \epsilon_t \tag{F.4}$$

where $\epsilon_t \sim N(0, 1)$. In this data generating process $\beta_{\text{true}} = 1$ is the benchmark for each bias test. For all values of y that are fixed, y is set to equal the most recent y value that could vary.

I then calculate the theoretical bias predicted by (F.3) for these three scenarios as well as estimate the empirical bias by regressing y on x regardless of y 's fixed status and subtracting the true β from my estimate. Finally, I also estimate the same regression, but omit observations with fixed y values. I repeat this exercise 1,000 times and calculate the mean and standard deviation of each statistic.

Table F.1 displays the results from these simulations. Each row presents the results for the varying levels of serial correlation in x . In columns 1 and 2, I present the mean and standard deviation of 1,000 simulations of the calculated theoretical and estimated bias respectively. column 3, presents the absolute value of the difference between each predicted and estimated bias. Column 4, presents the estimated bias for the regression that omits observations with fixed y values.

As expected, when the x values are perfectly serially correlated, there is no bias from estimating on the full or restricted samples. However, with strong, yet imperfect positive serial correlation, I predict about a 10 percent attenuation that is confirmed in the actual estimation. When x values have no serial correlation, the bias matches the extent to which y values are fixed, a 66. $\bar{7}$ percent attenuation in this case. Across each specification, the bias is completely mitigated by regressing only on years for which outcomes can vary. In my setting, $\rho_x(g) \approx 0.9$ and $\delta \approx 0.66\bar{7}$ meaning that naive estimates on annual data are theoretically predicted to be attenuated by about 7 percent.

Table F.1: Simulations to test Mechanical Bias

$\rho_x(3)$	District-Year Observations			Negotiaion Years Only
	Theoretical Bias (1)	Estimated Bias (2)	Abs(Difference) (3)	Estimated Bias (4)
1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
0.85	-0.097 (0.009)	-0.097 (0.005)	0.006 (0.005)	-0.000 (0.007)
0	-0.667 (0.009)	-0.667 (0.008)	0.003 (0.003)	0.000 (0.001)

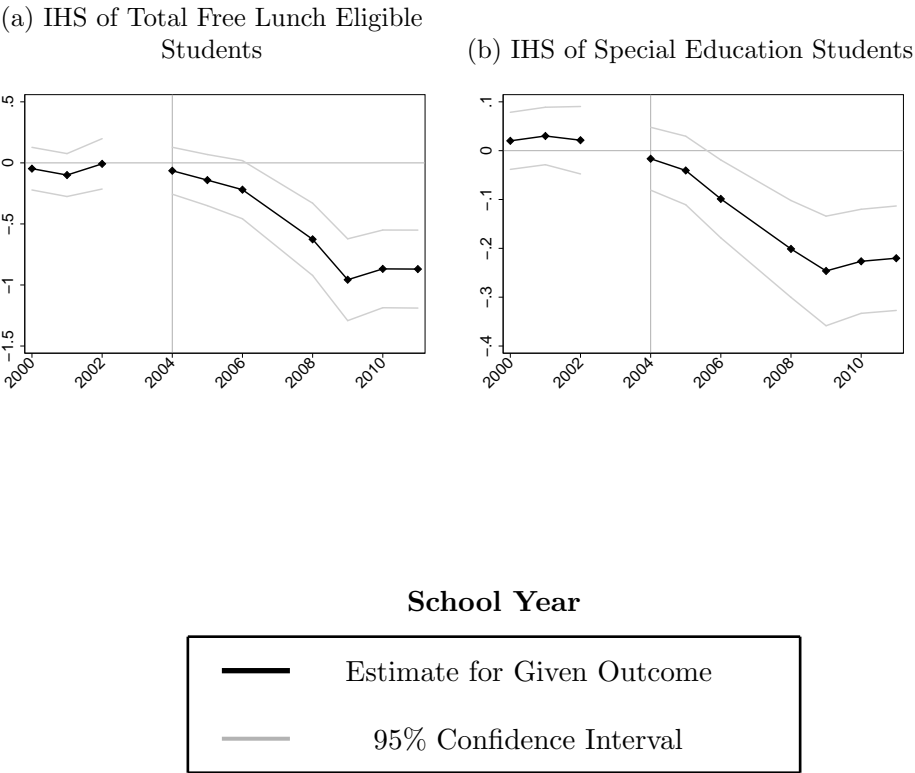
Notes: N= 10,000 simulated observations. $\delta = 0.667$. The theoretical bias, estimated bias, and absolute value difference in biases are calculated 1,000 times for the given autocorrelation values and the mean and standard deviations of these simulations are reported in columns 1-3. In all regressions, the true parameter value was unity. For $\rho_x(3) = 1$, x is simply an array of sequential positive integers. For $\rho_x(3) = .85$, x follows an AR(1) process that yields the given serial correlation. For $\rho_x(3) = 0$, $x \sim U[0, 100]$. For all regressions, y values are calculated as $y = 10 + 1 \cdot x + \epsilon$, where $\epsilon \sim N(0, 1)$. Column 4 presents the estimated bias for regressions run only on observations for which the outcome can vary (i.e., contract negotiation years).

G Unobservable Trend IV Check: All Outcomes

Figure G.1

Unobservable Trends Robustness Check: Student Mobility, Table 2.3

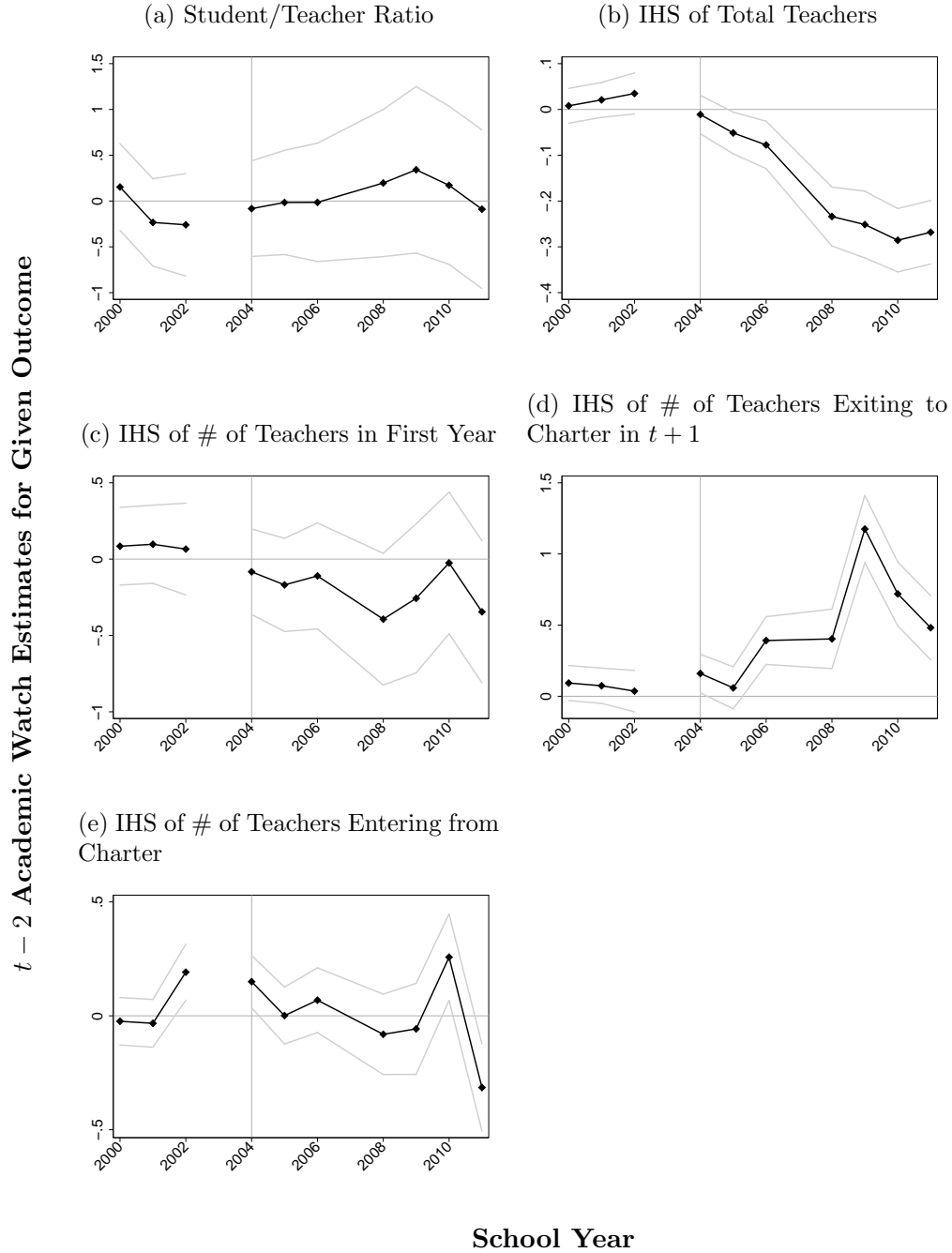
$t - 2$ Academic Watch Estimates for Given Outcome



Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.2

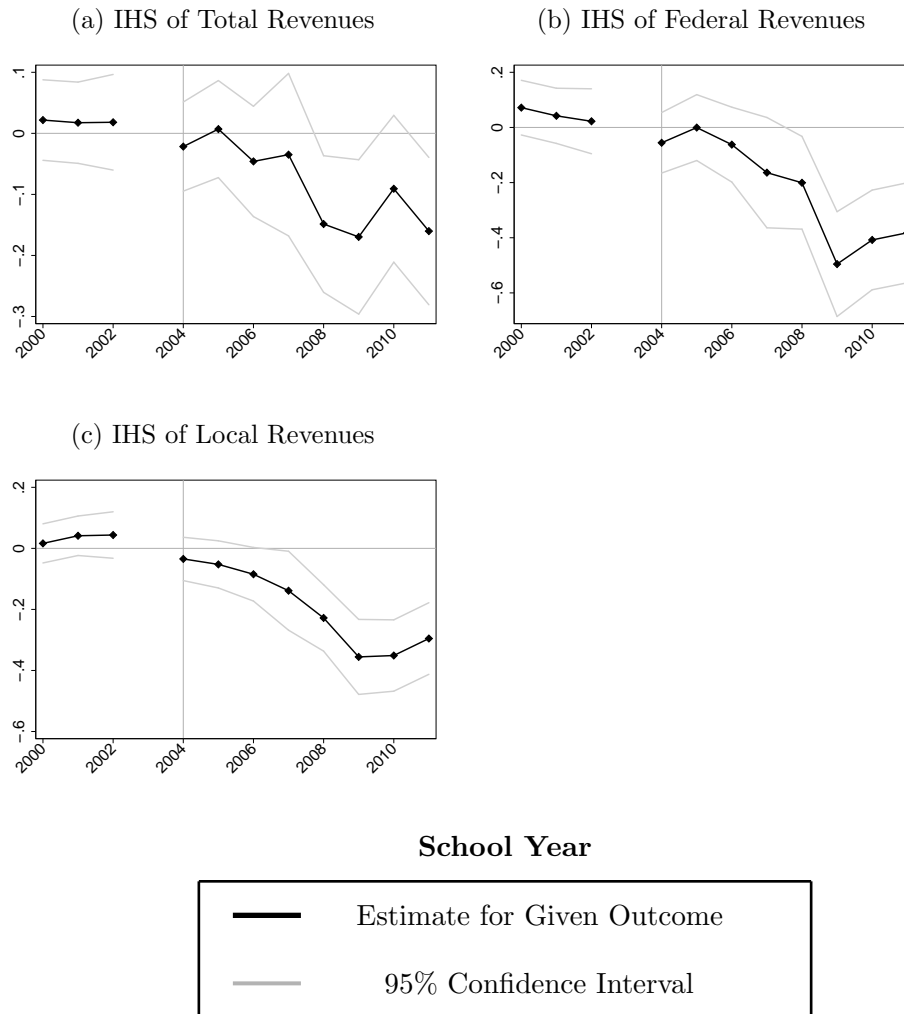
Unobservable Trends Robustness Check: Teacher Mobility, Table 2.3



Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.3

Unobservable Trends Robustness Check: District Revenues, Table 2.4 Panel A

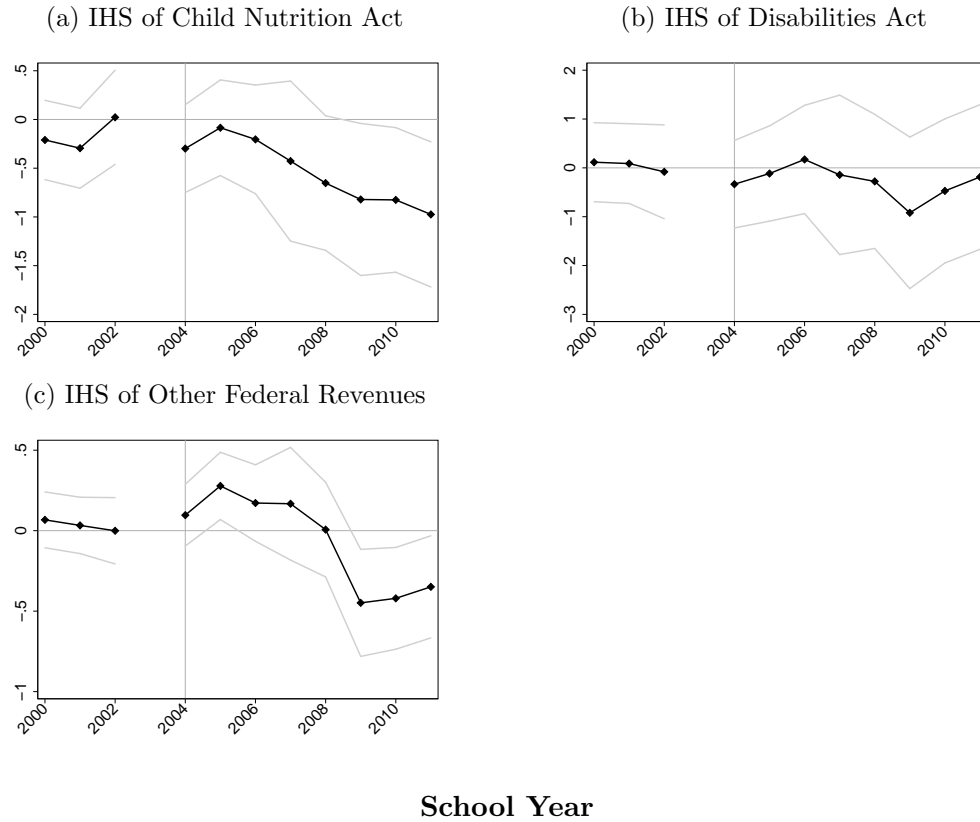


Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.4

Unobservable Trends Robustness Check: Federal Revenues, Table 2.4 Panel B

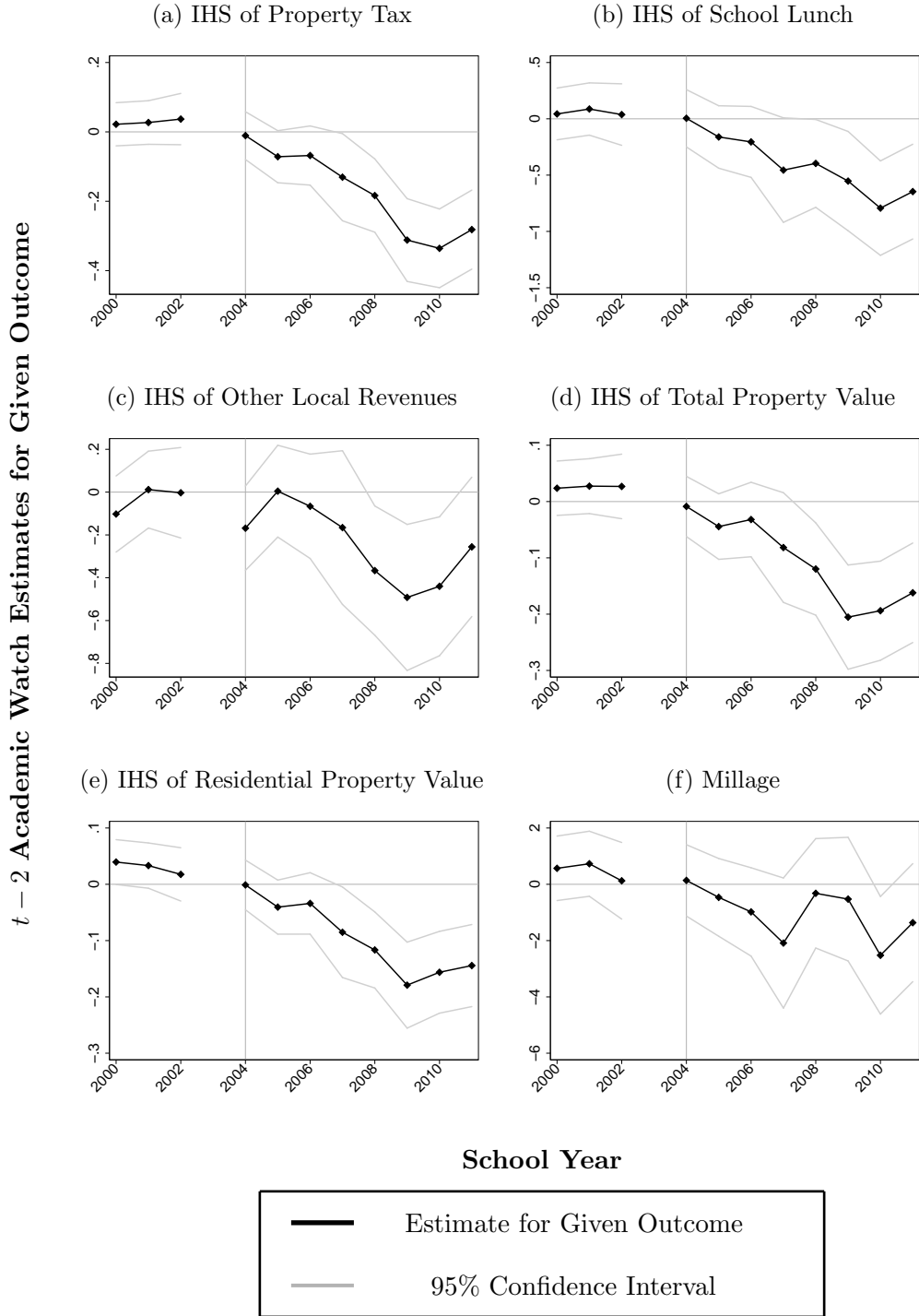
$t - 2$ Academic Watch Estimates for Given Outcome



Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.5

Unobservable Trends Robustness Check: Local Revenues, Table 2.4 Panels C & D

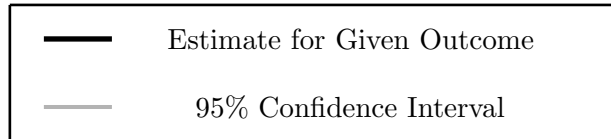
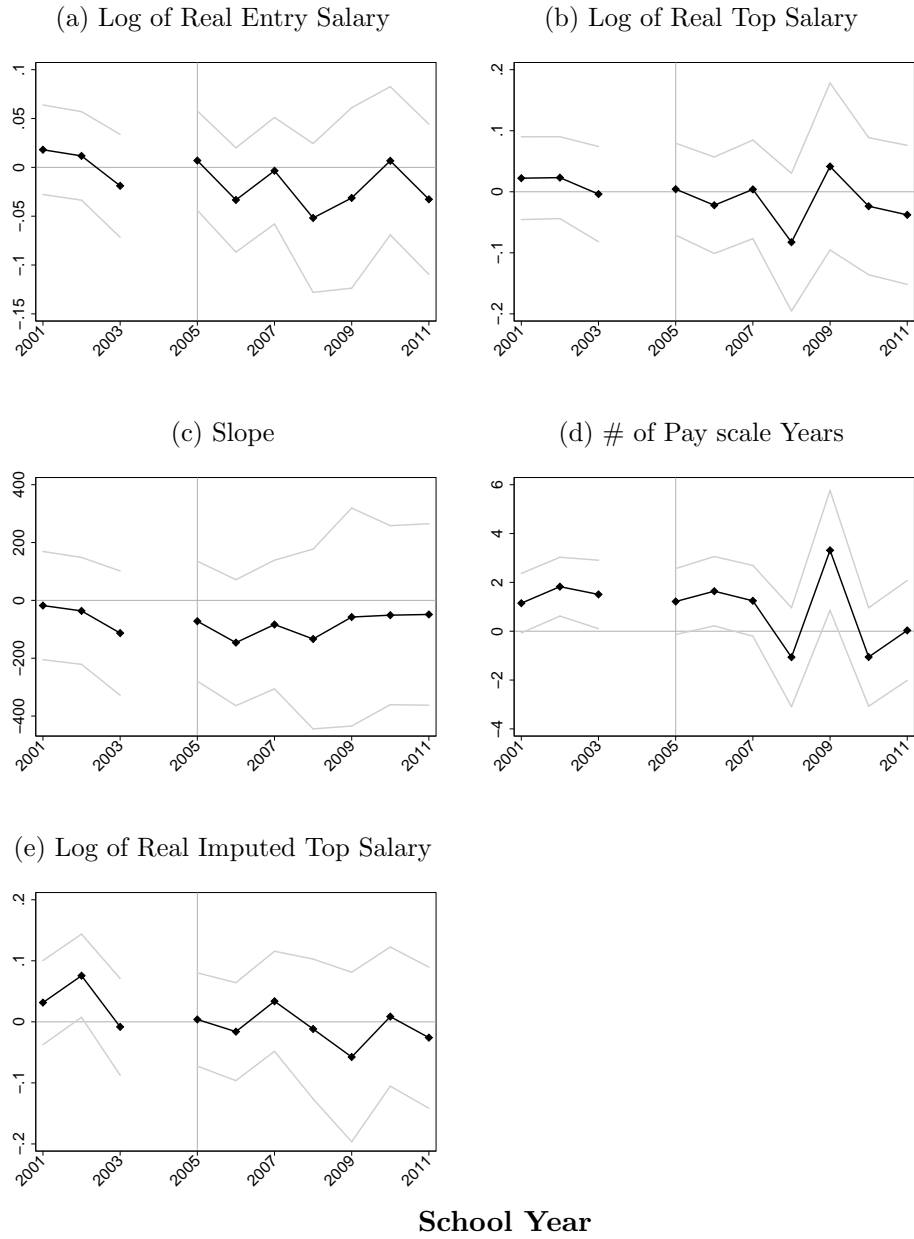


Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.6

Unobservable Trends Robustness Check: Contract Outcomes, Table 2.5

$t - 3$ Academic Watch Estimates for Given Outcome



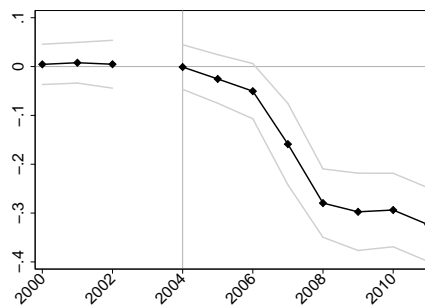
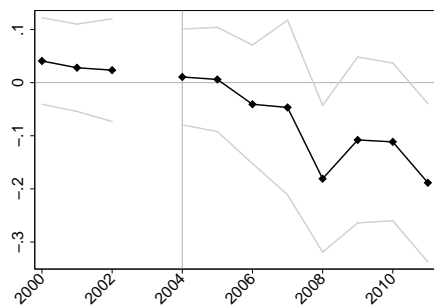
Notes: Each figure presents the effect of receiving an “Academic Watch” rating three years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

Figure G.7

Unobservable Trends Robustness Check: District Expenditures, Table 2.6

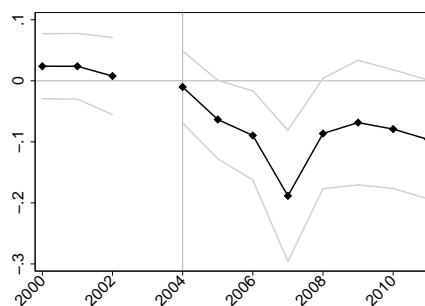
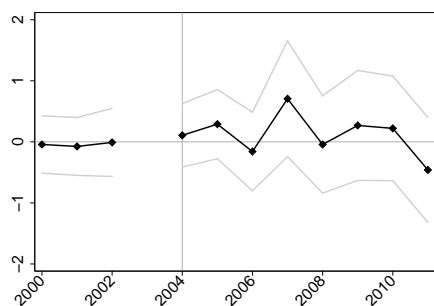
(a) IHS of Total Expenditures (Excluding Charter Payments)

(b) IHS of Instruction



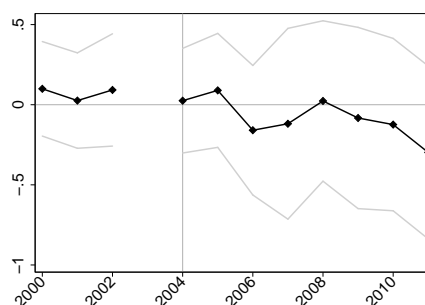
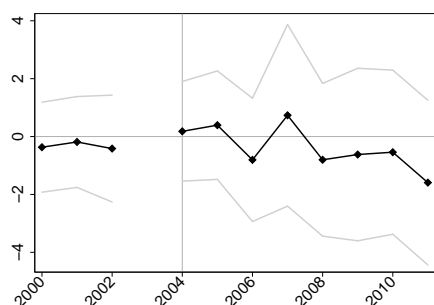
(c) IHS of Capital Outlays

(d) IHS of Other Expenditures

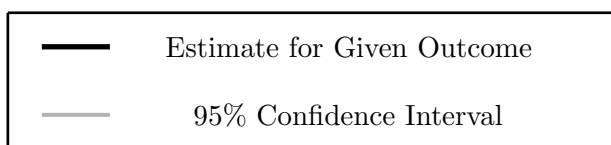


(e) IHS of New Construction

(f) IHS of Other Capital Outlays



School Year



Notes: Each figure presents the effect of receiving an “Academic Watch” rating two years earlier on the given current outcome, estimated from (2.8) as explained in Section 2.3.4. Each regression is respectively run on the sample restrictions for the given outcome in Sections 2.5 through 2.7.

H Regression Overidentification Tests and Outcome Means

Table H.1: Overidentification Tests: Hansen J Statistic

Panel A: <i>Student and Teacher Mobility</i>							
Potential Student Enrollment	Total FRL Eligible Students	Special Education Students	Stu/Tch Ratio	Total Teachers	# Tch. in First Year	# Tch. Exiting to CS	# Tch. Entering from CS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
8.995**	0.373	0.914	0.185	0.543	5.504*	0.986	4.728*
Panel B: <i>Federal District Revenues</i>							
Total	Child Nutrition Act	Disabilities Act	Other				
(1)	(2)	(3)	(4)				
3.479	0.861	0.105	1.997				
Panel C: <i>Local District Revenues</i>							
Total	Property Tax	School Lunch	Other	IHS of Total Property Value	IHS of Residential Property Value	Millage	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
2.173	3.613	1.268	0.240	0.233	0.360	3.480	
Panel D: <i>Collectively Bargained Outcomes</i>							
Entry Salary	Top Salary	Imputed Top Salary	Slope	# Payscale Years			
(1)	(2)	(3)	(4)	(5)			
3.923	9.925***	1.932	1.625	6.163**			
Panel E: <i>District Expenditures</i>							
Real Charter Payments (in 100,000s)	Total (Excluding CS Payment)	Instruction	Capital Outlays	Other	New Construction	Other	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
9.974***	0.545	7.566**	0.522	4.869*	0.438	2.151	

Notes: Each Panel and column provides the Hansen J overidentification test statistic for the respective regression samples found in Tables 2.3 through 2.6. ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

Table H.2: Table of Means: Dependent Variables

Panel A: <i>Student and Teacher Mobility</i>							
Potential Student Enrollment	Total FRL Eligible Students	Special Education Students	Stu/Tch Ratio	Total Teachers	# Tch. in First Year	# Tch. Exiting to CS	# Tch. Entering from CS
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2,897	763	396	17	176	13	0.136	0.100
Panel B: <i>Federal District Revenues</i>							
Total	Child Nutrition Act	Disabilities Act	Other				
(1)	(2)	(3)	(4)				
2,152,932	417,772	446,004	1,289,156				
Panel C: <i>Local District Revenues</i>							
Total	Property Tax	School Lunch	Other	IHS of Total Property Value	IHS of Residential Property Value	Millage	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
16,142,809	13,161,744	518,759	2,462,307	407,999,512	250,736,793	29	
Panel D: <i>Collectively Bargained Outcomes</i>							
Entry Salary	Top Salary	Imputed Top Salary	Slope	# Payscale Years			
(1)	(2)	(3)	(4)	(5)			
35,144	58,306	59,910	1,530	15			
Panel E: <i>District Expenditures</i>							
Real Charter Payments (in 100,000s)	Total (Excluding CS Payment)	Instruction	Capital Outlays	Other	New Construction	Other	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
0.462	31,027,948	15,649,987	3,289,005	12,088,956	2,499,978	789,028	
Panel F: <i>Charter Competition</i>							
Fraction of Students Transferring to Charter							
(1)							
0.010 (0.021)							

Notes: Each Panel and column provides the untransformed mean of the dependent variable for the respective regression samples found in Tables 2.3 through 2.6. Mean and standard deviation (in parentheses) for charter competition is provided in Panel F.

Chapter 6

Teacher Experience Appendix

A Identification of Experience Profiles

To see how identification of the returns to both general and three dimensions of context-specific experience might be secured, consider a simple case in which there are only two subjects, chemistry (C) and physics (P), and only two difficulty levels, basic (B) and honors (H). Suppose that four different teachers (not necessarily at the same school) each teach different subject-level combinations in their first years: Teacher 1 teaches basic physics (BP) in her first year, while teacher 2 teaches honors physics (HP), teacher 3 teaches basic chemistry (BC) and teacher 4 teaches honors chemistry (HC). Suppose then that all four teach honors chemistry (HC) every year thereafter. To keep the example simple, suppose further that the gains from each of the components of experience are fully persistent (no depreciation), and that each teacher only teaches classes in one subject-level per year. Panel A of Table F.1 displays the course assignment paths taken by each teacher, along with the observed stocks of general, subject-specific, level-specific, and subject-level specific experience that teachers will possess at the beginning of each of their school years.

Consider a difference-in-difference estimator that compares the change in teacher 1's average student test scores between years 2 and 3 with the corresponding change for teacher 2. Since each teacher teaches the same subject-level (HC) in both year 2 and year 3, focusing on changes over time differences out the permanent general and context-specific skills of the two teachers (as well as any differences in time-invariant school quality). Furthermore, comparing across teachers removes the common gains from the second year of (previous) general experience and the first year of subject-specific and subject-level specific experience. Because teacher 2 taught at the honors level in her first year, the extent to which teacher 1's performance converges to or diverges from teacher 2's performance between years 2 and 3 will reflect the relative value of the 2nd year of level-specific experience compared to the 1st year: $(d^l(2) - d^l(1)) - (d^l(1) - 0)$.¹ If instead we

¹Note that since returns to experience can only be identified relative to other levels of experience, we must

compare the change in student performance between years 3 and 4 for the same two teachers (1 and 2), we recover the relative value of the 3rd year of level-specific experience compared to the 2nd year: $(d^l(3) - d^l(2)) - (d^l(2) - d^l(1))$. Indeed, conditional on knowing the value of the first year of level-specific experience, $d^l(1)$, we can trace out the entire path of returns to level-specific experience by comparing the divergence/convergence in the performance of teachers 1 and 2 as they progress through their careers. If we replace teacher 2 with teacher 3 in the comparisons above, we instead trace out the path of returns to subject-specific experience. Now that the returns to subject-specific and level-specific experience have been identified, replacing teacher 3 with teacher 4 identifies the path of returns to subject-level-specific experience. Finally, the growth path of teacher 4, who never switched subjects or levels, identifies the path of returns to general experience.

To see how the value of the first year of experience might be identified for each component of experience, consider a second scenario in which teacher 1 teaches the following sequence of courses in her first four years: $BC \rightarrow HC \rightarrow BP \rightarrow HC$. Teacher 2 teaches the same set of courses, but in a different sequence: $BP \rightarrow HC \rightarrow BC \rightarrow HC$. Panel B of Table F.1 illustrates the stocks of general and context-specific experience each teacher possesses at the beginning of each year of teaching. Since both teachers teach honors chemistry with the same accumulated experience profile in year 4, comparing the performance of the two teachers identifies the difference in permanent teaching skill between the two teachers (part of which may be honors-chemistry specific): $\mu_{2CH} - \mu_{1CH}$. Once relative permanent skill has been identified, comparing the same two teachers' average student residuals in year 2 (when both were teaching honors chemistry) identifies the return to the 1st year of subject-specific experience, $d^j(1)$. Replacing basic chemistry with honors physics in this example would instead identify the return to the 1st year of level-specific experience ($d^l(1)$), while replacing it with honors chemistry would identify the return to the 1st year of subject-level specific experience ($d^{jl}(1)$). The return to the first year of general experience ($d^t(1)$) can then be identified via the growth in student average residuals from the 1st to the 2nd year from teachers who teach the same subject-level in each of their first two years.

While the sample histories used in these scenarios are stylized, note that there are many alternative moments that also provide identifying variation. Indeed, given the frequency with which subject and level switching occurs, we frequently observe multiple teachers who have taught the same set of subjects and levels over their careers at the school, but have taught them in different orders, or in different proportions. Since each different sequence also implies a different pattern of potential depreciation for a given model of depreciation, such comparisons allow us to simultaneously estimate the rates at which different experience components depreciate.²

Furthermore, each subject or level switch, regardless of the point in the career, provides a further source of identifying variation for the various context-specific experience profiles. Consequently, not only are these experience profiles estimable with reasonable precision (at least for the first several years of experience), but there are myriad overidentifying tests that can be implemented if one worries that particular sequences may be likely to occur in conjunction with particular changes in unobserved inputs (in violation of Assumption 1). Indeed, in Section 3.6 we show that the function linking four-dimensional stocks of general and context-specific teacher experience to

normalize one value for each function. We do so by setting $d^k(0) = 0$ for $k \in \{gen, j, l, lk\}$.

²In practice, after some experimentation, we include in our estimated specifications four dummy variables indicating whether the teacher taught the current subject last year, the current level last year, the current subject-level last year, and whether the teacher taught any class last year.

student performance is non-parametrically identified, and we present estimates from a more flexible (though noisily estimated) specification.

B Identification of Permanent Teaching Skill

To illustrate how $\hat{\mu}_{srjl}$ can be identified given any of Assumptions 2A-2C paired with 3A-3C, consider a teacher r' who teaches subject j' and level l' in school s' during years t_1 to t_2 . Let $Z_{ct} = Y_{ct} - X_{ct}\beta$ represent the average test score residual in classroom c at time t , after removing the component predictable based on classroom inputs. Then the average residual performance of students taught by teacher r' in school-subject-level combination (s', j', l') is given by:

$$\begin{aligned} E[Z_{ct}|(s, r, j, l) = (s', r', j', l')] \\ = \delta_{s'j'l'} + \mu_{s'r'j'l'} + \sum_{t'=t_1}^{t_2} w_{t'}[d^{gen}(exp_{r't'}^{gen}) + d^j(exp_{r't'}^j) + d^l(exp_{r't'}^l) + d^{jl}(exp_{r't'}^{jl})] \end{aligned} \quad (B.1)$$

where the weight $w_{t'}$ captures the fraction of all the students teacher r' taught in combination (s', j', l') that were taught in year t' . Since the experience profiles $d^{gen}(\cdot)$, $d^j(\cdot)$, $d^l(\cdot)$, and $d^{jl}(\cdot)$ were identified using comparisons of changes in performance across years in Section A, the average level of performance of teacher r' while teaching in school-subject level combination (s', j', l') identifies $\delta_{s'j'l'} + \mu_{s'r'j'l'}$. Under Assumption 2C, $\delta_{sjl} = 0 \forall (s, j, l)$, so this moment identifies $\mu_{s'r'j'l'}$ directly. Under Assumption 2A, we can use the fact that the (student weighted) average teacher quality in each school-subject-level is assumed to be zero. Specifically, the average residual performance of students in a particular school-subject-level is given by:

$$\begin{aligned} E[Z_{ct}|(s, j, l) = (s', j', l')] \\ = \delta_{s'j'l'} + E[d^{gen}(exp^{gen}) + d^j(exp^j) + d^l(exp^l) + d^{jl}(exp^{jl})|(s, j, l) = (s', j', l')], \end{aligned} \quad (B.2)$$

which identifies $\delta_{s'j'l'}$, leaving the teacher-specific average to identify $\mu_{s'r'j'l'}$. To identify $\delta_{s'}$ under Assumption 2B, we simply average at the school level instead of the school-subject-level level. Thus, μ_{srjl} can be identified for each combination of school-teacher-subject-level that we actually observe in the data.

C Recovering the Latent Variance Decomposition

This section shows how to distill the true decomposition of time-invariant skill into general, subject-specific, level-specific, and subject-level specific components from the estimated cell fixed effects $\{\hat{\mu}_{srjl}\}$. We first assume that each estimated school-teacher-subject-level fixed effect $\hat{\mu}_{srjl}$ can be written as the sum of the teacher's true context-specific skill and an uncorrelated error

component: $\hat{\mu}_{srjl} = \mu_{srjl} + \xi_{srjl}$. Let \mathcal{C} and C represent the set of classrooms and the total number of classrooms in the sample, respectively. In addition, let $\mu_{srjl(c)}$ represent the context-specific skill of the teacher that taught classroom c . Then the (student-weighted) sample variance in estimated context-specific skill can be decomposed as:

$$\frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\hat{\mu}_{srjl(c)})^2 = \frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\mu_{srjl(c)})^2 + \frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\xi_{srjl(c)})^2 \quad (\text{C.1})$$

where each w_c is a weight capturing the fraction of all student test scores in the sample that were associated with classroom c .

One would like to estimate the student-weighted variance in true teacher quality as:

$$\hat{Var}(\mu_{srjl}) = \frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\hat{\mu}_{srjl(c)})^2 - \frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\xi_{srjl(c)})^2. \quad (\text{C.2})$$

The sampling error components $\{\xi_{srjl}\}$ are not observed, but

$$\frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\xi_{srjl(c)})^2 \approx \frac{1}{C} \sum_{c \in \mathcal{C}} w_c E[(\xi_{srjl(c)})^2] = \frac{1}{C} \sum_{c \in \mathcal{C}} w_c se(\xi_{srjl(c)})^2, \quad (\text{C.3})$$

so we estimate the error variance component using the standard error estimates for each school-teacher-subject-level fixed effect:

$$\hat{Var}(\mu_{srjl}) = \frac{1}{C} \sum_{c \in \mathcal{C}} w_c (\hat{\mu}_{srjl(c)})^2 - \frac{1}{C} \sum_{c \in \mathcal{C}} w_c se(\xi_{srjl(c)})^2. \quad (\text{C.4})$$

By using the delta method to estimate standard errors for $\hat{\mu}_{srjl}$, denoted $se(\tilde{\mu}_{srjl})$, we can estimate $\hat{Var}(\tilde{\mu}_{srjl})$ analogously. Then, $\hat{Var}(\bar{\mu}_{sr})$ can be estimated via:

$$\hat{Var}(\bar{\mu}_{sr}) = \hat{Var}(\mu_{srjl}) - \hat{Var}(\tilde{\mu}_{srjl}) \quad (\text{C.5})$$

To prevent teachers who only taught a single subject-level combination from biasing our estimate of $\hat{Var}(\bar{\mu}_{sr})$ downward, when calculating $\hat{Var}(\tilde{\mu}_{srjl})$ we restrict the sample of school-teacher-subject-level combinations to those in which the relevant school-teacher combination was observed in at least two school-teacher-subject-level combinations.

Further use of the delta method allows the same procedure to be applied in recovering the true variance of subject-specific, level-specific, and subject-level-specific teacher talent.³

³Specifically, we calculate the true variances as follows. First, consider the alternative decomposition $\tilde{\mu}_{srjl} = \bar{\mu}_{srj} + (\tilde{\mu}_{srjl} - \bar{\mu}_{srj})$. We estimate the true variance of the second component by first using the delta method to calculate standard errors for $(\tilde{\mu}_{srjl} - \bar{\mu}_{srj})$ and then applying the same method as above. We then obtain the variance

D Testing the Additive Separability of Context-Specific Experience Profiles

D.1 Smoothing the Nonparametric Experience Contribution Function

To address the volatility of our experience cell fixed effect estimates, we assume that the true structural function $d(*, *, *, *)$ is differentiable everywhere, and smooth our estimates using a kernel function featuring the normal PDF with zero mean and standard deviation 0.5 (denoted ϕ despite the non-unity standard deviation):

$$\tilde{d}(\mathbf{exp}) = \frac{\sum_{\mathbf{exp}' \in \mathcal{EX}} w_{\mathbf{exp}'} \phi(|\mathbf{exp}' - \mathbf{exp}|) \hat{d}(\mathbf{exp}')}{\sum_{\mathbf{exp}' \in \mathcal{EX}} w_{\mathbf{exp}'} \phi(|\mathbf{exp}' - \mathbf{exp}|)}, \quad (\text{D.1})$$

where $\hat{d}(\mathbf{exp}')$ is the estimate on the given experience profile from equation (3.13). The argument to the normal density $|\mathbf{exp}' - \mathbf{exp}|$ is the L1 norm or taxicab distance between the two experience profiles: $|\mathbf{exp}' - \mathbf{exp}| = |exp^{gen'} - exp^{gen}| + |exp^{j'} - exp^j| + |exp^{l'} - exp^l| + |exp^{sl'} - exp^{sl}|$. The weight $w_{\mathbf{exp}'}$ represents the fraction of observations in the sample in which the experience profile exp' is observed. Thus, the impact of $\hat{d}(1, 1, 1, 1)$ on $\tilde{d}(1, 1, 1, 0)$ will be greater than that of $\hat{d}(1, 1, 0, 0)$, despite equal L1 distances, because $\hat{d}(1, 1, 1, 1)$ is much more precisely estimated than $\hat{d}(1, 1, 0, 0)$. The chosen bandwidth yields a four-dimensional function $\tilde{d}(*, *, *, *)$ that is smooth enough to remove considerable sampling error, yet is still flexible enough to reveal true complementarities where they may occur.

D.2 Marginal Effects Example

This subsection shows how we estimate profiles of returns to single dimensions of experience from the smoothed nonparametric experience cell estimates. These profiles can then be compared with the corresponding dimension-specific profiles from our additively-separable baseline specification. For each initial value v of each component of experience, we approximate the average partial derivative of the experience production function at v ($E[\frac{\partial d(exp^{gen}, exp^j, exp^l, exp^{sl})}{\partial exp^{dim}} | exp^{dim} = v]$ for $dim \in \{gen, j, l, sl\}$) by calculating a weighted average marginal effect of an extra year of the chosen component of experience (holding the other experience components fixed). The weighted average is taken over all combinations of the other three experience dimensions that are observed among experience cells that feature the chosen initial value v in the chosen dimension $dim \in \{gen, j, l, sl\}$.

For example, let $\mathcal{Q}^{j,v}$ denote the set of experience cells at which a partial derivative for

in subject-specific teaching talent, $\hat{Var}(\bar{\mu}_{srj})$, via $\hat{Var}(\bar{\mu}_{srj}) = \hat{Var}(\tilde{\mu}_{srjl}) - \hat{Var}((\tilde{\mu}_{srjl} - \bar{\mu}_{srj}))$. The variance in level-specific teaching talent, $\hat{Var}(\bar{\mu}_{srl})$, can be calculated using an identical approach. Finally, we estimate the variance in subject-level-specific teaching talent using: $\hat{Var}(\bar{\mu}_{srjl} - \bar{\mu}_{srj} - \bar{\mu}_{srl}) = \hat{Var}(\tilde{\mu}_{srjl}) - \hat{Var}(\bar{\mu}_{srj}) - \hat{Var}(\bar{\mu}_{srl})$.

subject-specific experience at initial value v may be calculated:

$$\begin{aligned} \mathcal{Q}^{j,v} &= \{(exp^t, exp^j, exp^l, exp^{sl}) : \\ &exp^j = v, (exp^t, exp^j, exp^l, exp^{sl}) \in \mathcal{D}, (exp^t, exp^j + 1, exp^l, exp^{sl}) \in \mathcal{D}\}. \end{aligned}$$

Then the average marginal effect of subject-specific experience among cells featuring $exp^j = v$ can be calculated via:

$$\begin{aligned} \frac{\bar{\partial} d(exp^t, v, exp^l, exp^{sl})}{\partial exp^j} &= \\ \sum_{k \in \mathcal{Q}^{j,v}} w_k [\hat{d}(exp_k^t, v + 1, exp_k^l, exp_k^{sl}) - \hat{d}(exp_k^t, v, exp_k^l, exp_k^{sl})] \end{aligned}$$

The weight w_k is composed of the product of two sub-weights associated with the two experience cells included in the partial derivative estimate. Each sub-weight represents the fraction of all teacher-school-subject-level-year cells that feature the chosen experience combination. The w_k are then re-scaled to sum to 1.

E Methodology for Measuring Forecast Bias

This appendix describes the implementation of the test for forecast bias in our estimates of teacher talent discussed in Section 3.6.3. The intuition for the test is that if the estimated fixed effects $\{\hat{\mu}_{srjl}\}$ properly capture the true talent contributions $\{\mu_{srjl}\}$, then differences in fixed effects among teachers in a chosen context estimated from one partition of our data should predict differences in mean residual achievement in the same context in a second, left out partition. Our methodology closely mirrors that of Chetty et al. (2014).

The first step is to construct an appropriate sample of school-teacher-subject-level (hereafter SRJL) combinations. Two conditions must be met for a given SRJL combination to enter the test sample. First, the teacher must have taught at least three classes in the chosen school-subject-level (hereafter SJJL). At least one class must be available for the forecasted partition, and at least two classes must be available for the partition that forms the basis for the forecast (so that a valid standard error for the estimate $\hat{\mu}_{srjl}$ can be formed).⁴ Second, the SJJL context associated with the chosen SRJL combination must be shared by at least one other

⁴If only one class is available for given SRJL combination for the partition within which production function is estimated, the estimated value $\hat{\mu}_{srjl}$ will be chosen to perfectly fit the classroom mean test score residual, and there will be no regression error with which to form heteroskedasticity-robust standard errors. While other approaches to estimating standard errors (estimating at the test score level, imposing homoskedastic standard errors) would not require this second classroom in the forecasting partition, we want the method for constructing standard errors in the model used for the forecast test to mimic as closely as possible the one employed for our main estimates.

teacher. This ensures that persistent school inputs that are specific to (or differentially important in) the chosen course-level can be differenced out. 2,285 out of an original 18,257 SRJL combinations satisfy these two restrictions.

Let $Z_{ct}^F \equiv Y_{ct} - X_{ct}\hat{\beta}_{jl} - \hat{d}^{gen}(exp_{rt}^{gen}) - \hat{d}^j(exp_{rt}^j) - \hat{d}^l(exp_{rt}^l) - \hat{d}^{jl}(exp_{rt}^{jl})$ represent classroom level residuals, where the values $\hat{\beta}_{jl}$, \hat{d}^{gen} , \hat{d}^j , \hat{d}^l , and \hat{d}^{jl} are used to form the residuals represent the estimated coefficients on the student demographics, year dummies, and context-specific experience profiles from our restricted specification (3.3).

We randomly select one classroom from each eligible SRJL to form the forecasted partition (Partition 2), and assign the remaining classrooms to Partition 1. Let $Z_{ct}^F \equiv [Z_{ct}^{1F}, Z_{ct}^{2F}]$ capture the corresponding partition of classroom-level residuals. We fit the following model to the class-level achievement data from Partition 1:

$$Z_{ct}^1 = \mu_{srjl}^F + \epsilon_{ct}^F \quad (\text{E.1})$$

Since the estimated fixed effects from this model, $\hat{\mu}_{srjl}^F$, still contain the contributions of school-subject-level inputs δ_{sjl} , we choose a teacher r' from each SJL environment and subtract this teacher's context-specific fixed effect from those of the other teachers in the SJL environment to form the differences $\hat{\mu}_{srjl}^F - \hat{\mu}_{sr'jl}^F$.

Because these differences still contain sampling error, the coefficient in a regression of differences in residuals $Z_{c(s,r,j,l)t}^{2F} - Z_{c(s,r',j,l)t}^{2F}$ from the forecasted partition on the estimated fixed effect differences $(\hat{\mu}_{srjl}^F - \hat{\mu}_{sr'jl}^F)$ will be attenuated toward zero even when these fixed effect differences are unbiased estimates of true differences in task-specific talent. Still following Chetty et al. (2014), we therefore shrink the estimated fixed effect differences by a pair-specific reliability ratio to form empirical-Bayes difference estimates: $Diff_{srjl}^{EB} = (\frac{Var(\mu_{srcl})}{Var(\mu_{srcl}) + (\hat{\sigma}_{srjl}^F)^2})(\hat{\mu}_{srjl}^F - \hat{\mu}_{sr'jl}^F)$, where $(\hat{\sigma}_{srjl}^F)^2$ is the squared standard error of the estimated fixed effect difference (obtained from the component fixed effect estimates via the delta method), and $Var(\mu_{srjl})$ is the estimated true variance in teacher talent contributions across classrooms (.154) from our lower bound decomposition presented in Section 3.5.1.

We then regress the vector of differences in mean classroom residuals in the forecasted sample (Partition 2) on the shrunk fixed effect differences $Diff_{srjl}^{EB}$:

$$Z_{c(s,r,j,l)t}^{2F} - Z_{c(s,r',j,l)t}^{2F} = \beta^F Diff_{srjl}^{EB} + e_{ct}^{2F} \quad (\text{E.2})$$

If the estimates of both the true variance in teacher talent contributions across classrooms (the “signal”) and the standard errors of the fixed effect differences (the “noise”) are correct, the coefficient β^F should converge in probability to 1, so that the talent estimates are “forecast unbiased” (Chetty et al. (2014)).

While this test captures the ability of our specification to consistently estimate the combined general and context-specific talent of a given teacher teaching in a given context, the ability to choose teachers' classroom assignments in a way that maximizes achievement contributions depends critically on the ability to isolate and consistently estimate only the context-specific components of teacher fixed contributions to achievement. Thus, we also construct two additional

forecast tests that capture the degree to which our estimates of subject-specific and level-specific talent can forecast out-of-sample subject-specific and level-specific comparative advantages, respectively.

Unfortunately, unlike our tests of the consistency of our combined talent estimates, which could be performed using differences among teachers who taught in the same SJL context, testing the consistency of comparative advantage estimates requires evaluating the degree to which difference-in-differences between teachers who taught the same two courses at the same school can be forecast. Thus, a given SRJL combination only enters the subject (level) forecast sample if (1) the teacher taught at least three classrooms in both the chosen SJL *and* a second course that shares the same school-level (school-subject) environment, and (2) there exists a second teacher who also taught at least three classrooms in the same two school-subject-level contexts. These criteria are far more stringent. Much of the variation that identified the estimated true variances in subject-specific and level-specific talent came from difference-in-differences in which at least one of the teachers taught fewer than three subjects in at least one of the school-subject-level contexts. Indeed, applying these criteria leaves us with 205 and 289 difference-in-differences on which to perform the forecast test for subject-specific and level-specific talent estimates, respectively.

The methodology for the context-specific forecast tests is otherwise perfectly analogous to the forecast test for combined teacher talent. Difference-in-differences in residual mean test scores from among the left-out classrooms in Partition 2 across teachers and either subjects or levels (conditioning on the same school-level or school-course environment as appropriate) are regressed on empirical Bayes estimates of difference-in-differences in teacher context-specific talent from the forecasting sample:

$$\begin{aligned} (Z_{c(s,r,j,l)t}^{2F} - Z_{c(s,r,j',l)t}^{2F}) - (Z_{c(s,r',j,l)t}^{2F} - Z_{c(s,r',j',l)t}^{2F}) &= \beta^F Diff_in_Diff_{srjl}^{EB} + e_{ct}^{F,j} \\ (Z_{c(s,r,j,l)t}^{2F} - Z_{c(s,r,j,l')t}^{2F}) - (Z_{c(s,r',j,l)t}^{2F} - Z_{c(s,r',j,l')t}^{2F}) &= \beta^F Diff_in_Diff_{srjl}^{EB} + e_{ct}^{F,l} \end{aligned} \quad (E.3)$$

The results of the test for forecast bias in the combined talent estimates as well as the corresponding tests for the subject-specific and level-specific talent estimates are presented in Table F.7.

F Formulation of the Counterfactual Simulation

To formulate the static problem, first let \mathcal{J} represent the set of subjects offered within a given school-field combination. Similarly, let \mathcal{L} represent the set of levels, and let \mathcal{JL} represent the set of subject-level combinations. Let C_{jl} represent the number of classes to be staffed in subject-level combination $jl \in \mathcal{JL}$, with $N_c = \sum_{jl \in \mathcal{JL}} C_{jl}$ denoting the total number of classes to be staffed. Let \mathcal{R} represent the set of teachers, with R elements. As before, exp_r^{gen} captures the number of prior years in which teacher r has taught any classroom, and exp_r^j , exp_r^l , and exp_r^{jl} capture the number of prior years in which teacher r has taught at least one classroom in subject j , level l , and subject-level combination jl , respectively. Student and classroom contributions $X_{ct}\beta_{jl}$ can be ignored, since they are assumed to be constant across counterfactual reallocations (and are assumed to be additively separable from teacher inputs).

Using (1) the estimated smoothed non-parametric experience production function from the “full” specification (3.13) introduced in Section 3.6.4, denoted $\hat{d}(exp_{rt}^{gen}, exp_{rt}^{j(c)}, exp_{rt}^{l(c)}, exp_{rt}^{jl(c)})$, and (2) the empirical Bayes estimated posterior belief about teacher r ’s context-specific talent for increasing test scores in school s in subject-level combination $(j(c), l(c))$, denoted $\mu_{srj(c)l(c)}^{EB}$,⁵ we can predict the contribution of context-specific experience to the counterfactual performance of teacher r in classroom c in school s in a given year t via:

$$\hat{Y}_{rt}^c = \mu_{srj(c)l(c)}^{EB} + \hat{d}(exp_{rt}^{gen}, exp_{rt}^{j(c)}, exp_{rt}^{l(c)}, exp_{rt}^{jl(c)}). \quad (\text{F.2})$$

The goal is to choose the mapping $f : \mathcal{C} \rightarrow \mathcal{R}$ from classrooms to teachers that maximizes the sum of student test scores, subject to the constraints that each teacher can only teach as many classrooms as he/she was observed teaching at time t (denoted \bar{C}_r), and every classroom must be taught by exactly one teacher⁶:

$$\begin{aligned} & \max_{f: \mathcal{C} \rightarrow \mathcal{R}} \sum_{c \in \mathcal{C}} \hat{Y}_{f(c)}^c \\ & s.t. \quad \sum_r \mathbb{1}(f(c) = r) = 1 \quad \forall c \\ & s.t. \quad \sum_c \mathbb{1}(f(c) = r) = \bar{C}_r \quad \forall r \end{aligned} \quad (\text{F.3})$$

where $\mathbb{1}(f(c) = r)$ indicates that teacher r is assigned to classroom c .

This optimization problem can be recast as a binary integer programming problem:

⁵ $\mu_{srj(c)l(c)}^{EB}$ is calculated by shrinking the fixed effect estimate $\hat{\mu}_{srj(c)l(c)}$ toward zero (the global mean contribution) by multiplying it by the reliability ratio:

$$\mu_{srjl}^{EB} = (\hat{\mu}_{srjl}) \left(\frac{Var(\mu_{srjl} - \bar{\mu}_{sr})}{\hat{\sigma}_{\hat{\mu}_{srjl}}^2 + Var(\mu_{srjl} - \bar{\mu}_{sr})} \right). \quad (\text{F.1})$$

$Var(\mu_{srjl} - \bar{\mu}_{sr})$ is the estimated true variance in subject-level deviations from school-teacher general talent taken from Row 3 of Column 1 of Table 3.4. $\hat{\sigma}_{\hat{\mu}_{srjl}}^2$ is the estimated squared standard error from the fixed effect estimate $\hat{\mu}_{srjl}$, which captures the contribution of noise (sampling error) to the variance in the school-teacher-subject-level fixed effect estimates $\{\mu_{srjl}\}$. We set $\mu_{srjl}^{EB} = 0$ for school-teacher-subject-level combinations that are feasible in the simulated assignment but are never observed in our sample.

⁶We suppress dependence on the year (t) in what follows.

$$\begin{aligned}
& \max_{\mathbf{x}} \mathbf{a} * \mathbf{x} \\
& s.t. \ M_c * \mathbf{x} = 1 \ \forall \ c \\
& s.t. \ N_r * \mathbf{x} = \overline{C}_r \ \forall \ r \\
& s.t. \ \mathbf{x} \in \{0, 1\}
\end{aligned} \tag{F.4}$$

\mathbf{a} consists of a $1 \times (C * R)$ row vector of predicted student performances for each potential teacher-classroom combination:

$$\mathbf{a} = \left(\hat{Y}_1^1 \quad \dots \quad \hat{Y}_1^C \quad \hat{Y}_2^1 \quad \dots \quad \hat{Y}_2^C \quad \dots \quad \hat{Y}_R^1 \quad \dots \quad \hat{Y}_R^C \right)$$

\mathbf{x} consists of a $(C * R) \times 1$ vector of potential teacher assignments:

$$\mathbf{x} = \begin{pmatrix} x_1^1 \\ \vdots \\ x_1^C \\ x_2^1 \\ \vdots \\ x_2^C \\ \vdots \\ x_R^1 \\ \vdots \\ x_R^C \end{pmatrix}$$

where $x_r^c = \mathbb{1}(f(c) = r)$ is an indicator for whether teacher r is assigned to classroom c .

M_c consists of a $1 \times C * R$ row vector capturing the number of teachers assigned to classroom c (restricted to be 1 $\forall \ c$):

$$M_c = \left(\underbrace{\overbrace{0 \dots 0}^{c-1} \ 1 \ \overbrace{0 \dots 0}^{C-c}}_{\text{repeated R times}} \dots \overbrace{0 \dots 0}^{c-1} \ 1 \ \overbrace{0 \dots 0}^{C-c} \right)$$

N_r consists of a $1 \times C * R$ row vector capturing the number of classrooms taught by teacher r (restricted to be equal to \bar{C}_r , the number taught in the sample):

$$N_r = \left(\overbrace{0 \dots 0}^{(r-1)*C} \underbrace{1 \dots 1}_C \overbrace{0 \dots 0}^{(R-r)*C} \right).$$

We solve this binary integer programming problem for each school-field combination in the first year of the sample. We then update each teacher’s context-specific experience profile for the second year given the experience they gained under the optimal assignment in the first year.⁷ We repeat this process until the end of the sample so as to reap the long-run rewards associated with accumulating high levels of relevant context-specific experience. The “static” version of the simulation does not update each teacher’s context-specific experience profile for the next year after allocating teachers in a given year, but instead treats every year in the sample as if it were the first year.

Some of our simulations exploit the full sample of teachers, rather than restricting attention to those teachers with fully observed teaching histories. For teachers who begin teaching after 1995, when our sample begins, we impute their teaching history as of 1995 by randomly assigning them the teaching history of a full history teacher who 1) was observed (later in the sample) at the 1995 general experience level of the imputed teacher, and 2) who shares the same most commonly taught subject-level across all the years of our sample as the imputed teacher. Some teachers are sufficiently experienced in 1995 so that there is no teacher with a fully observed teaching history who is ever observed at such a high level of general experience in our sample. These teachers are randomly assigned a 1995 teaching history from among the full history teachers who are observed at 12+ years of general experience who share the same most commonly taught subject-level. Once a 1995 teaching history has been imputed for all teachers with missing histories, we accumulate their post-1995 stocks of general and context-specific experience as it existed in the data (if constructing actual stocks) or as it was optimally assigned (if constructing simulated stocks).

Finally, some of our simulation results tables use a random allocation of teachers to classrooms as a baseline rather than the actual allocation of teachers. To ensure that the random allocations are feasible, we construct them by selecting, for each school-field, a random permutation of the allocation identified as the solution to the binary assignment problem.

⁷Since non-tested subjects are not reallocated, any general or level-specific experience teachers accumulated in those subjects under the true allocation is also included in the update.

Table F.1: Identification Example: Experience Stocks for Hypothetical Teachers in Each Year

Panel A: Identifying Variation for Experience Profiles, Example 1										
Year	Teacher 1: New Subj/Lvl					Teacher 2: New Subj Only				
	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BP	0	0	0	0	HP	0	0	0	0
2	HC	1	0	0	0	HC	1	0	1	0
3	HC	2	1	1	1	HC	2	1	2	1
4	HC	3	2	2	2	HC	3	2	3	2
5	HC	4	3	3	3	HC	4	3	4	3

Year	Teacher 3: New Lvl Only					Teacher 4: Same Subj/Lvl				
	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BC	0	0	0	0	HC	0	0	0	0
2	HC	1	1	0	0	HC	1	1	1	1
3	HC	2	2	1	1	HC	2	2	2	2
4	HC	3	3	2	2	HC	3	3	3	3
5	HC	4	4	3	3	HC	4	4	4	4

Panel B: Identifying Variation for Experience Profiles, Example 2										
Year	Teacher 1					Teacher 2				
	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.	Crs.	Gen.	Subj.	Lvl.	Subj.-Lvl.
1	BC	0	0	0	0	BP	0	0	0	0
2	HC	1	1	0	0	HC	1	0	0	0
3	BP	2	0	1	0	BC	2	1	1	0
4	HC	3	2	1	1	HC	3	2	1	1

Notes: This table provides the path of experience stocks for each teacher in each of the two examples illustrating experience profile identification that occur in Appendix Section A. Each entry provides the level of general or task-specific experience in the dimension indicated by the column heading at the beginning of the year associated with the row. “B”-Basic, “H”-Honors, “P”-Physics, “C”-Chemistry.

Table F.2: The Distribution of Years of Experience among Classes Taught by 2nd and 3rd Year Teachers

<i>Panel A: Second-Year Teachers</i>				
General	Subject	Level	Subj.-Lvl	%
1	1	1	1	70.8%
1	1	1	0	2.7%
1	1	0	0	4.5%
1	0	1	0	19.4%
1	0	0	0	2.5%

<i>Panel B: Third-Year Teachers</i>				
General	Subject	Level	Subj.-Lvl	%
2	2	2	2	54.9%
2	2	2	1	3.0%
2	2	2	0	0.6%
2	2	1	1	4.1%
2	2	1	0	0.6%
2	2	0	0	2.2%
2	1	2	1	17.7%
2	1	2	0	1.0%
2	1	1	1	1.7%
2	1	1	0	0.8%
2	1	0	0	1.1%
2	0	2	0	10.5%
2	0	1	0	0.8%
2	0	0	0	1.1%

Notes: The table presents the classroom-weighted distribution of four-dimensional experience stocks among second- and third-year teachers in our final sample. 10,270 and 8,665 total classes were taught by second-year and third-year teachers respectively. Note that multiple subject-level combinations can be taught in a year.

Table F.3: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Baseline Specification)

	(1)
# of Concurrent Classes Taught	0.000 [0.000]
# of Concurrent Subject-Level Combinations Taught	0.000 [0.002]
1(Did Not Teach Last Year)	0.004 [0.020]
1(Did Not Teach Subject Last Year)	-0.003 [0.013]
1(Did Not Teach Level Last Year)	0.006 [0.015]
1(Did Not Teach Subject-Level Last Year)	-0.001 [0.012]
1(Did Not Teach in Last 2 Years)	-0.006 [0.036]
1(Did Not Teach Subject in Last 2 Years)	-0.014 [0.026]
1(Did Not Teach Level in Last 2 Years)	0.003 [0.028]
1(Did Not Teach Subject-Level in Last 2 Years)	0.002 [0.022]
1(Final Year Teaching)	-0.005 [0.019]
1(Final Year Teaching Subject)	0.001 [0.012]
1(Final Year Teaching Level)	-0.004 [0.018]
1(Final Year Teaching Subject-Level)	-0.010 [0.011]

Notes: Regression also contains a full set of school-subject-level and school-teacher-subject-level fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table 3.5 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.4: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Restricted Specification)

	(1)
# of Concurrent Classes Taught	-0.000 [0.000]
# of Concurrent Subject-Level Combinations Taught	-0.004* [0.002]
1(Did Not Teach Last Year)	0.014 [0.017]
1(Did Not Teach Subject Last Year)	-0.020* [0.011]
1(Did Not Teach Level Last Year)	0.004 [0.012]
1(Did Not Teach Subject-Level Last Year)	0.005 [0.010]
1(Did Not Teach in Last 2 Years)	0.005 [0.031]
1(Did Not Teach Subject in Last 2 Years)	-0.014 [0.019]
1(Did Not Teach Level in Last 2 Years)	-0.001 [0.022]
1(Did Not Teach Subject-Level in Last 2 Years)	0.000 [0.017]
1(Final Year Teaching)	-0.004 [0.014]
1(Final Year Teaching Subject)	-0.029*** [0.008]
1(Final Year Teaching Level)	0.008 [0.012]
1(Final Year Teaching Subject-Level)	-0.017** [0.007]

Notes: Regression also contains a full set of school-subject-level and school-teacher fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table 3.6 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.5: Tests for Dynamic Course Assignment Responses to Unobserved Time-Varying Endogenous Inputs

Year Relative to Change	Permanent Teacher Changes				Temporary Teacher Changes			
	General (1)	Subject (2)	Level (3)	SL (4)	General (5)	Subject (6)	Level (7)	SL (8)
$t - 1$	0.010 [0.008]	0.005 [0.007]	0.039** [0.022]	-0.016 [0.016]	-0.000 [0.001]	-0.001 [0.001]	0.000 [0.001]	0.001 [0.001]
$t - 2$	-0.012 [0.012]	-0.016* [0.011]	-0.021 [0.017]	0.022 [0.019]	0.001 [0.002]	0.003* [0.002]	0.001 [0.002]	0.002 [0.002]
$t - 3$	-0.003 [0.012]	0.032** [0.016]	0.037** [0.016]	0.035* [0.024]	0.005** [0.003]	0.003 [0.003]	-0.001 [0.003]	-0.001 [0.003]
$t - 4$	0.017* [0.013]	-0.010 [0.014]	0.014 [0.038]	-0.009 [0.017]	0.002 [0.004]	0.001 [0.003]	0.006* [0.004]	0.004 [0.004]
$t - 5$	-0.003 [0.029]	-0.014 [0.013]	0.017 [0.049]	0.024 [0.037]	-0.006* [0.005]	-0.002 [0.004]	0.002 [0.004]	0.004 [0.005]
$t - 6$	-0.021 [0.018]	0.002 [0.019]	-0.083*** [0.031]	0.032 [0.067]	0.008* [0.006]	0.005 [0.005]	-0.001 [0.006]	0.005 [0.006]
$t - 7$	-0.040 [0.038]	-0.012 [0.022]	-0.182 [0.160]	-0.115** [0.055]	0.005 [0.008]	-0.001 [0.005]	0.015*** [0.006]	-0.006 [0.007]

Notes: Table entries display average school-teacher-year residuals (Columns 1 and 5), school-teacher-subject-year residuals (Columns 2 and 6), school-teacher-level-year residuals (Columns 3 and 7), and school-teacher-subject-level-year residuals (Columns 4 and 8), respectively, in the years leading up to a change in classroom assignment (using residuals from the *Restricted Specification* in equation (3.3)). A permanent change in general course assignment (Column 1) is defined as a teacher-year combination in which the teacher is not observed teaching any course in a subsequent sample year. A permanent change in subject assignment (Column 2) is defined as a teacher-subject-year combination in which the teacher teaches the chosen subject, but is not observed teaching the chosen subject again in subsequent sample years. Permanent changes in level (Column 3) and subject-level (Column 4) assignments are defined analogously to permanent subject changes. Temporary changes in assignment (Columns 5-8) are defined in a similar manner as permanent course assignment changes, except the teacher is observed returning to teach (Column 5) or teach in the chosen subject (Column 6), level (Column 7), or subject-level (Column 8) in a subsequent sample year. Bootstrap standard errors (in brackets) are computed using 1,000 iterations. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively.

Table F.6: Backcasting Test for Non-Random Student Sorting (Restricted Specification with Classroom Average 7th Grade Math Scores as the Outcome Variable)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	-0.027*** [0.009]	-0.005 [0.007]	0.024*** [0.009]	0.002 [0.006]	-0.005* [0.004]
2 yrs	-0.014* [0.010]	-0.022*** [0.009]	0.016** [0.010]	0.016** [0.008]	-0.004 [0.004]
3 yrs	-0.026** [0.013]	-0.021** [0.011]	0.032*** [0.011]	0.024*** [0.010]	0.009** [0.005]
4 yrs	-0.019* [0.013]	-0.030*** [0.011]	0.022** [0.012]	0.025*** [0.010]	-0.002 [0.005]
5-6 yrs	-0.016 [0.014]	-0.020* [0.013]	0.020* [0.013]	0.026** [0.012]	0.010** [0.005]
7-10 yrs	0.008 [0.015]	-0.028** [0.015]	-0.009 [0.014]	0.039*** [0.014]	0.011* [0.006]
11-14 yrs	0.007 [0.020]	0.009 [0.025]	-0.011 [0.020]	0.012 [0.026]	0.017* [0.013]

Notes: $N = 61,993$ test-score-weighted classroom observations. Results are based on an altered version of the *Restricted Specification* in equation (3.3) in which the actual classroom average of students current test scores from the chosen class are replaced by the classroom average of the 7th grade math scores of these students. Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.7: Testing for Forecast Bias in Estimates of Time-Invariant Task-Specific Teacher Talent

<u>Forecasting Sample</u>	<u>Outcome (Forecasted Sample)</u>		
	Diff (Sch-Tea-Subj-Lvl)	Diff-in-Diff (Subj)	Diff-in-Diff (Lvl)
	(1)	(2)	(3)
Diff ^{EB} (Sch-Tea-Subj-Lvl)	0.825		
$[\hat{\mu}_{srjl} - \hat{\mu}_{sr'jl}]^{EB}$	(0.019)		
Diff-in-Diff (Subj)		1.013	
$[(\hat{\mu}_{srjl} - \hat{\mu}_{srjl'}) - (\hat{\mu}_{sr'jl} - \hat{\mu}_{sr'jl'})]^{EB}$		(0.242)	
Diff-in-Diff (Lvl)			0.456
$[(\hat{\mu}_{srjl} - \hat{\mu}_{srjl'}) - (\hat{\mu}_{sr'jl} - \hat{\mu}_{sr'jl'})]^{EB}$			(0.333)
Observations	7,246	205	289

Notes: The entries in this table are coefficients (with standard errors in brackets) capturing the degree of forecast bias in estimates of combined (general and task-specific) talent, subject-specific talent, and level-specific talent, respectively, from a set of split sample tests. See Appendix Section E for a detailed description of the methodology. In each specification, the estimator should yield a forecast coefficient that converges in probability to 1 if our achievement production function is correctly specified. Specifically, the outcome in Column 1 is the difference in average test score residuals among a pair of classes from the same school-subject-level taught by two different teachers from a partition of our main sample. The entry in Row (1), Column (1) captures the coefficient on a vector of empirical Bayes forecasts of the expected difference in achievement among these pairs of teachers based on (appropriately shrunk) estimates of the difference in their combined general and course-specific productivity from a second, mutually exclusive partition used to construct the forecast. The entries in Column 2 and Column 3 replace these pair-specific differences on both sides of the equation with differences-in-differences among pairs of teachers across common pairs of courses that differ only in subject (Column 2) or level (Column 3). These coefficients capture the degree of forecast bias in the model's ability to estimate subject-specific and level-specific comparative advantages. Heteroskedasticity-robust (White) standard errors (in brackets) are computed for each coefficient.

Table F.8: Effect of Number of Courses of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification)

Course Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 crs	0.019 [0.036]	0.011 [0.012]	-0.006 [0.012]	0.008 [0.009]	0.031 [0.035]
2 crs	0.051** [0.027]	0.007 [0.011]	0.002 [0.011]	0.002 [0.009]	0.062*** [0.025]
3 crs	0.075*** [0.017]	0.021** [0.011]	-0.005 [0.012]	0.002 [0.010]	0.094*** [0.015]
4-5 crs	0.066*** [0.013]	0.032*** [0.011]	0.001 [0.011]	0.004 [0.010]	0.104*** [0.008]
6-9 crs	0.054*** [0.012]	0.049*** [0.012]	0.009 [0.012]	0.002 [0.011]	0.114*** [0.006]
10-20 crs	0.070*** [0.014]	0.062*** [0.013]	0.005 [0.013]	-0.006 [0.013]	0.131*** [0.007]
21+ crs	0.082*** [0.015]	0.071*** [0.015]	0.003 [0.015]	-0.009 [0.016]	0.146*** [0.010]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3.3). Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the number of classes taught in prior years by the classroom's teacher in total (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.9: True Variances in Fixed Effects (Using Course-Based Measure of Teacher Experience with the Baseline Specification)

	Lower Bound		Intermediate		Upper Bound	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0237	0.154	0.0467	0.216	0.0604	0.246
General Talent	0.0176	0.133	0.0368	0.192	0.0505	0.225
Subj-Lvl Combos	0.0061	0.078	0.0099	0.099	0.0099	0.099
Sch-Lvl-Tch Combos	0.0198	0.141	0.0407	0.202	0.0544	0.233
Subject Talent	0.0039	0.063	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0216	0.147	0.0433	0.208	0.0569	0.239
Level Talent	0.0021	0.045	0.0034	0.059	0.0034	0.059
Subject-Level Talent	0.0001	0.011	0.0005	0.022	0.0005	0.022

Notes: This variance decomposition is based on a version of the baseline specification (equation 3.2 in which experience in each context dimension (general, subject-specific, level-specific, and subject-level-specific) is measured as the total number of previous classrooms taught in the relevant context. *Lower Bound* estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.3.2 for details.

Table F.10: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with the Sample Restricted to Classrooms Featuring Teachers in their First School)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.062*** [0.012]	0.018** [0.011]	-0.003 [0.011]	0.011 [0.009]	0.088*** [0.005]
2 yrs	0.087*** [0.016]	0.022** [0.014]	-0.011 [0.014]	0.018* [0.012]	0.117*** [0.006]
3 yrs	0.093*** [0.018]	0.035** [0.016]	-0.016 [0.016]	0.014 [0.014]	0.126*** [0.007]
4 yrs	0.105*** [0.020]	0.035** [0.017]	-0.022 [0.018]	0.012 [0.015]	0.130*** [0.008]
5-6 yrs	0.109*** [0.021]	0.034** [0.020]	-0.019 [0.019]	0.023* [0.017]	0.147*** [0.010]
7-10 yrs	0.116*** [0.025]	0.020 [0.025]	-0.031* [0.023]	0.027 [0.023]	0.131*** [0.013]
11-14 yrs	0.101*** [0.034]	0.019 [0.044]	0.016 [0.035]	-0.014 [0.049]	0.122*** [0.033]

Notes: $N = 51,773$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3.3). The sample is restricted to classrooms featuring a teacher that is teaching in his/her first school (i.e. the teacher's entire teaching history was acquired at the current school). Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.11: True Variances in Fixed Effects (Using the Baseline Specification with the Year-Based Measure of Teacher Experience and a Sample Restricted to Classrooms Featuring Teachers in Their First Schools)

	Lower Bound		Intermediate		Upper Bound	
	Var.	SD	Var.	SD	Var.	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Sch-Subj-Lvl-Tch Combos	0.0225	0.150	0.0454	0.213	0.0590	0.243
General Talent	0.0167	0.129	0.0356	0.189	0.0491	0.222
Subj-Lvl Combos	0.0058	0.076	0.0098	0.099	0.0098	0.099
Sch-Lvl-Tch Combos	0.0188	0.137	0.0395	0.199	0.0530	0.230
Subject Talent	0.0037	0.061	0.0060	0.077	0.0060	0.077
Sch-Subj-Tch Combos	0.0205	0.143	0.0421	0.205	0.0556	0.236
Level Talent	0.0019	0.044	0.0033	0.058	0.0033	0.058
Subject-Level Talent	0.0001	0.011	0.0005	0.023	0.0005	0.023

Notes: *Lower Bound* estimates allocate all of the between school-subject-level variance in residual test scores to school and student inputs (Assumption 2A). This is implemented by including school-subject-level fixed effects and normalizing the mean among school-teacher-subject-level fixed effects to be 0 in each school-subject-level. *Intermediate* estimates allocate the between school variance in residual test scores to school and student inputs, and the within-school/between subject-level variance to teachers (Assumption 2B). This is implemented by replacing the school-subject-level fixed effects with school fixed effects only. *Upper Bound* estimates allocate all of the between school-subject-level variance in residual test scores to teachers (Assumption 2C). This is implemented by removing all school-level controls. See Section 3.3.2 for details.

Table F.12: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with Linear Depreciation)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.066*** [0.011]	0.013* [0.009]	-0.004 [0.010]	0.015** [0.008]	0.089*** [0.004]
2 yrs	0.086*** [0.014]	0.023** [0.012]	-0.005 [0.012]	0.014* [0.010]	0.118*** [0.006]
3 yrs	0.095*** [0.016]	0.036*** [0.014]	-0.006 [0.014]	0.006 [0.012]	0.131*** [0.007]
4 yrs	0.103*** [0.018]	0.040*** [0.015]	-0.009 [0.016]	0.004 [0.014]	0.137*** [0.008]
5-6 yrs	0.106*** [0.019]	0.040** [0.018]	-0.000 [0.017]	0.005 [0.015]	0.151*** [0.009]
7-10 yrs	0.117*** [0.022]	0.025 [0.021]	-0.004 [0.020]	-0.001 [0.019]	0.136*** [0.012]
11-14 yrs	0.111*** [0.028]	0.027 [0.038]	0.031 [0.028]	-0.029 [0.041]	0.140*** [0.026]

Notes: Regression specification mimics the *Restricted Specification* (see Equation (3.3) and the notes to Table 3.6), except that the indicator sets for whether the teacher failed to teach a course in each of the relevant dimensions of context (general, subject, level, and subject-level) last year or in the last two years are replaced by linear controls for the number of years since the teacher taught any course and since the teacher taught in the subject, level, and subject-level associated with the classroom observation. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.13: Coefficient Estimates Associated with Control Variables Capturing Teacher Workload, Depreciation of Experience Capital, and Productivity Declines in the Last Year Teaching Any Class or Teaching a Class in the Chosen Subject, Level, or Subject-Level Combination (Restricted Specification with Linear Depreciation)

	(1)
# of Concurrent Classes Taught	-0.000 [0.000]
# of Concurrent Subject-Level Combinations Taught	-0.003 [0.002]
# of Years Since Last Taught	0.004 [0.006]
# of Years Since Last Taught Subject	-0.005 [0.004]
# of Years Since Last Taught Level	0.001 [0.004]
# of Years Since Last Taught Subject-Level	-0.005 [0.004]
1(Final Year Teaching)	-0.005 [0.014]
1(Final Year Teaching Subject)	-0.028*** [0.008]
1(Final Year Teaching Level)	0.008 [0.012]
1(Final Year Teaching Subject-Level)	-0.018** [0.007]

Notes: Regression specification mimics the *Restricted Specification* (see Equation (3.3) and the notes to Table 3.6), except that the indicator sets for whether the teacher failed to teach a course in each of the relevant dimensions of context (general, subject, level, and subject-level) last year or in the last two years are replaced by linear controls for the number of years since the teacher taught any course and since the teacher taught in the subject, level, and subject-level associated with the classroom observation. Regression also contains a full set of school-subject-level and school-teacher fixed effects, calendar year effects, a set of observable classroom covariates, and a set of four additively separable flexibly parameterized profiles capturing productivity gains from years of general, subject-specific, level-specific, and subject-level-specific experience. See Table F.12 for estimates of these experience profiles. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.14: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with 7th Grade Math and Reading Test Scores Added as Controls)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.069*** [0.011]	0.014* [0.009]	-0.007 [0.010]	0.014** [0.008]	0.090*** [0.004]
2 yrs	0.088*** [0.014]	0.025** [0.012]	-0.007 [0.012]	0.013* [0.010]	0.119*** [0.006]
3 yrs	0.099*** [0.016]	0.037*** [0.014]	-0.013 [0.014]	0.007 [0.012]	0.131*** [0.007]
4 yrs	0.106*** [0.018]	0.042*** [0.015]	-0.016 [0.015]	0.005 [0.014]	0.138*** [0.008]
5-6 yrs	0.108*** [0.019]	0.042*** [0.017]	-0.006 [0.017]	0.007 [0.015]	0.151*** [0.009]
7-10 yrs	0.115*** [0.022]	0.026 [0.021]	-0.009 [0.019]	0.002 [0.019]	0.135*** [0.011]
11-14 yrs	0.108*** [0.028]	0.018 [0.037]	0.025 [0.028]	-0.014 [0.041]	0.137*** [0.026]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3.3). Refer to notes below Table 3.5 for a full description of the control variables. This regression also includes 7th grade math and reading standardized test scores as additional controls. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every classroom of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.15: Heterogeneity across Subject Fields in the Effects of Years of General and Subject-Specific Experience on Student Test Scores (Restricted Specification with Level and Subject-Level Experience Additionally Constrained to 0)

Years Exp.	Math		Science		Social Studies		English	
	General	Subject	General	Subject	General	Subject	General	Subject
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 yr	0.072*** [0.012]	0.021** [0.010]	0.070*** [0.013]	0.046*** [0.011]	0.072*** [0.014]	0.017 [0.012]	0.020 [0.016]	0.013 [0.015]
2 yrs	0.090*** [0.014]	0.038*** [0.012]	0.098*** [0.017]	0.047*** [0.015]	0.091*** [0.017]	0.027* [0.015]	0.017 [0.022]	0.035* [0.021]
3 yrs	0.102*** [0.016]	0.048*** [0.014]	0.101*** [0.019]	0.062*** [0.018]	0.091*** [0.018]	0.032* [0.017]	0.020 [0.026]	0.036 [0.026]
4 yrs	0.095*** [0.017]	0.053*** [0.015]	0.089*** [0.022]	0.078*** [0.021]	0.119*** [0.020]	0.026 [0.019]	0.036 [0.031]	0.022 [0.031]
5-6 yrs	0.120*** [0.018]	0.057*** [0.017]	0.089*** [0.025]	0.075*** [0.024]	0.119*** [0.021]	0.042** [0.022]	0.023 [0.037]	0.025 [0.037]
7-10 yrs	0.134*** [0.020]	0.038* [0.020]	0.062** [0.032]	0.075** [0.031]	0.133*** [0.022]	0.005 [0.026]	0.038 [0.045]	0.009 [0.045]
11-14 yrs	0.154*** [0.028]	0.008 [0.042]	0.073 [0.047]	0.097* [0.051]	0.125*** [0.036]	-0.081 [0.058]	0.066 [0.055]	0.042 [0.058]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. Results are based on an altered version of the *Restricted Specification* in equation (3.3) in which 1) we impose the additional restrictions that gains from level-specific and subject-level-specific experience are constrained to be 0: $d^l(exp) = 0$ and $d^{jl}(exp) = 0 \forall exp$, and 2) we generalize the gains from years of general and subject-specific experience to be field-specific: $d^{gen}(exp) \rightarrow d_{field}^{gen}(exp)$, $d^j(exp) \rightarrow d_{field}^j(exp)$, $field \in \{\text{Math, Science, Social Studies, English}\}$. Refer to notes below Table 3.5 for a full description of the control variables. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (odd Columns) or in the subject (even Columns) associated with the current classroom observation. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.16: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification with Full Set of Indicator Variables for Each Observed Years of Experience)

Years Experience	General	Subject	Level	Subj.-Level	Combined
	(1)	(2)	(3)	(4)	(5)
1 yr	0.064*** [0.011]	0.013* [0.009]	-0.002 [0.010]	0.013** [0.008]	0.089*** [0.004]
2 yrs	0.084*** [0.014]	0.021** [0.012]	-0.003 [0.012]	0.016* [0.010]	0.118*** [0.006]
3 yrs	0.093*** [0.016]	0.033*** [0.014]	-0.005 [0.014]	0.010 [0.012]	0.131*** [0.007]
4 yrs	0.100*** [0.018]	0.036*** [0.015]	-0.008 [0.016]	0.010 [0.014]	0.137*** [0.008]
5 yrs	0.101*** [0.019]	0.041*** [0.017]	0.003 [0.017]	0.012 [0.016]	0.156*** [0.009]
6 yrs	0.101*** [0.021]	0.021 [0.019]	-0.001 [0.018]	0.026* [0.018]	0.146*** [0.011]
7 yrs	0.117*** [0.023]	0.018 [0.022]	-0.017 [0.020]	0.019 [0.020]	0.137*** [0.012]
8 yrs	0.110*** [0.026]	-0.017 [0.026]	0.002 [0.023]	0.028 [0.024]	0.123*** [0.014]
9 yrs	0.120*** [0.028]	-0.001 [0.031]	0.002 [0.026]	0.027 [0.029]	0.148*** [0.017]
10 yrs	0.120*** [0.032]	-0.020 [0.036]	0.006 [0.029]	0.021 [0.037]	0.127*** [0.022]
11 yrs	0.104*** [0.035]	0.006 [0.047]	0.056** [0.032]	-0.025 [0.050]	0.140*** [0.029]
12 yrs	0.153*** [0.044]	-0.086* [0.062]	-0.014 [0.042]	0.147** [0.065]	0.199*** [0.039]
13 yrs	0.087* [0.053]	-0.063 [0.105]	0.010 [0.055]	0.176* [0.113]	0.210*** [0.044]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. Refer to notes below Table 3.5 for a full description of the control variables. Results are based on an altered version of the *Restricted Specification* from equation (3.3) in which bins for years of experience 5-6, 7-11, and 11+, respectively, are replaced by indicator variables for each individual year of experience (general, subject, level, and subject-level combination). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Column 5, entitled *Combined*, captures the combined predicted contribution of all four dimensions of experience capital for the case in which the teacher has taught the course associated with the classroom observation in every year of a career length defined by the row label. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.17: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification Featuring Quartics in Each Dimension of Context-Specific Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
Year Exp.	0.063*** [0.012]	0.019* [0.011]	0.006 [0.010]	0.007 [0.010]
(Year Exp.) ²	-0.014*** [0.004]	-0.003 [0.004]	-0.004 [0.003]	-0.001 [0.004]
(Year Exp.) ³	0.001*** [0.000]	0.000 [0.001]	0.001 [0.000]	-0.000 [0.001]
(Year Exp.) ⁴	-0.000*** [0.000]	-0.000 [0.000]	-0.000 [0.000]	0.000 [0.000]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3.3). Refer to notes below Table 3.5 for a full description of the control variables. Experience profiles in this regression are generated by replacing year-of-experience dummy variables from the restricted specification with a quartic in each of the four dimensions of experience (general, subject, level, and subject-level). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.18: Effect of Years of General, Subject-Specific, Level-Specific, and Subject-Level-Specific Experience on Student Test Scores (Restricted Specification Featuring Quartics in Each Dimension of Context-Specific Experience: Predicted Values for First 10 Years of Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
Pred. Exp. in Years 1	0.050*** [0.009]	0.016** [0.008]	0.003 [0.008]	0.007 [0.007]
Pred. Exp. in Years 2	0.078*** [0.013]	0.027** [0.011]	0.001 [0.012]	0.011 [0.010]
Pred. Exp. in Years 3	0.092*** [0.016]	0.032** [0.013]	-0.002 [0.014]	0.014 [0.011]
Pred. Exp. in Years 4	0.098*** [0.017]	0.034** [0.015]	-0.005 [0.015]	0.015 [0.013]
Pred. Exp. in Years 5	0.099*** [0.018]	0.032* [0.016]	-0.006 [0.016]	0.014 [0.014]
Pred. Exp. in Years 6	0.100*** [0.020]	0.027 [0.018]	-0.005 [0.017]	0.013 [0.016]
Pred. Exp. in Years 7	0.101*** [0.022]	0.019 [0.021]	-0.002 [0.019]	0.013 [0.019]
Pred. Exp. in Years 8	0.105*** [0.024]	0.009 [0.025]	0.004 [0.021]	0.014 [0.022]
Pred. Exp. in Years 9	0.110*** [0.027]	-0.004 [0.029]	0.010 [0.024]	0.019 [0.027]
Pred. Exp. in Years 10	0.116*** [0.030]	-0.018 [0.034]	0.016 [0.027]	0.030 [0.032]

Notes: $N = 61,993$ test-score-weighted classroom observations. The outcome is the class average of student standardized test scores in the subject. *Restricted Specification* refers to equation (3.3). Refer to notes below Table 3.5 for a full description of the control variables. Experience profiles in this regression are generated by replacing year-of-experience dummy variables from the restricted specification with a quartic in each of the four dimensions of experience (general, subject, level, and subject-level). Predicted values are used for the first 10 years of experience in each dimension. Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience estimates from Table F.17, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See Section 3.2 for methodological details.

Table F.19: Average Accumulated Marginal Effects Derived from Non-Parametric Experience Production Function (Full Specification with Year-Based Definition of Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
1 yr	0.021*** [0.007]	0.024** [0.013]	0.006 [0.012]	0.014** [0.007]
2 yrs	0.073** [0.033]	0.043** [0.019]	0.006 [0.018]	0.016* [0.012]
3 yrs	0.076** [0.044]	0.080*** [0.026]	0.016 [0.023]	0.016 [0.017]
4 yrs	0.073** [0.044]	0.094*** [0.032]	0.030 [0.028]	0.019 [0.021]
5-6 yrs	0.097** [0.047]	0.098*** [0.039]	0.051** [0.031]	0.018 [0.024]
7-10 yrs	0.114** [0.050]	0.117*** [0.049]	0.035 [0.034]	0.016 [0.028]
11-14 yrs	0.128*** [0.053]	0.135*** [0.056]	0.072** [0.041]	0.007 [0.037]

Notes: Refer to the notes below Table 3.5 for a full description of the control variables. Experience profiles are generated by integrating partial derivatives of extra experience in each experience dimension, evaluated at each level of experience, over all the levels of experience. These partial derivatives are derived from a smoothed version of the non-parametrically estimated production function for experience gains described in equation (3.13). Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience-cell fixed effects, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See D.2 for methodological details.

Table F.20: Average Accumulated Marginal Effects Derived from Non-Parametric Experience Production Function (Restricted Specification with Year-Based Definition of Experience)

Years Experience	General	Subject	Level	Subj.-Level
	(1)	(2)	(3)	(4)
1 yr	0.017*** [0.004]	0.029*** [0.007]	0.020*** [0.006]	0.013*** [0.004]
2 yrs	0.062*** [0.016]	0.035*** [0.010]	0.029*** [0.009]	0.017*** [0.005]
3 yrs	0.052** [0.027]	0.058*** [0.013]	0.025** [0.011]	0.015** [0.008]
4 yrs	0.041* [0.027]	0.067*** [0.016]	0.020** [0.012]	0.014* [0.010]
5-6 yrs	0.096*** [0.027]	0.065*** [0.020]	0.025** [0.014]	0.015* [0.011]
7-10 yrs	0.092*** [0.031]	0.065*** [0.025]	0.011 [0.016]	0.011 [0.014]
11-14 yrs	0.116*** [0.036]	0.069** [0.031]	0.059*** [0.020]	0.005 [0.027]

Notes: Refer to the notes below Table 3.6 for a full description of the control variables. Experience profiles are generated by integrating partial derivatives of extra experience in each experience dimension, evaluated at each level of experience, over all the levels of experience. These partial derivatives are derived from a smoothed version of the non-parametrically estimated production function for experience gains described in equation (3.13), but where the school-teacher-subject-level fixed effects μ_{srjl} are restricted to be common across subject-levels within a school-teacher combination: $\mu_{srjl} = \bar{\mu}_{st} \forall (j, l)$ and (s, r) . Experience is measured as the total number of prior years in which the classroom's teacher taught at least one class at all (Col. 1) or in the subject (Col. 2), level (Col. 3), or subject-level (Col. 4) associated with the current classroom observation. Standard errors (in brackets) are calculated by applying the delta method to the cluster-robust standard errors for the experience-cell fixed effects, which were clustered at the teacher level. Significance at the 1%, 5%, and 10% levels are represented by ***, **, and * respectively. See D.2 for methodological details.

Table F.21: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations Separately by Field (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

Eligible Teach.		Math				Science				Social Studies			
		Static		Dynamic		Static		Dynamic		Static		Dynamic	
		Actual (1)	Random (2)	Actual (3)	Random (4)	Actual (5)	Random (6)	Actual (7)	Random (8)	Actual (9)	Random (10)	Actual (11)	Random (12)
2	Total	.015	.018	.016	.021	.017	.026	.017	.028	.018	.027	.019	.028
	Talent	.015	.015	.015	.017	.019	.021	.020	.022	.020	.022	.020	.022
	Exper.	-.000	.003	.000	.004	-.002	.005	-.003	.005	-.002	.004	-.001	.005
3	Total	.029	.035	.030	.041	.024	.040	.026	.041	.025	.034	.027	.038
	Talent	.029	.031	.030	.034	.026	.033	.027	.033	.026	.029	.027	.029
	Exper.	-.000	.004	.000	.007	-.002	.006	-.001	.008	-.001	.005	.000	.008
4	Total	.033	.043	.035	.045	.030	.043	.032	.044	.031	.038	.033	.044
	Talent	.034	.039	.034	.036	.032	.034	.032	.033	.032	.032	.032	.034
	Exper.	-.001	.004	.001	.009	-.002	.009	-.000	.011	-.001	.006	.001	.010
5-6	Total	.041	.049	.043	.051	.034	.050	.037	.052	.038	.046	.040	.053
	Talent	.041	.044	.041	.042	.036	.039	.037	.039	.038	.039	.038	.042
	Exper.	-.000	.005	.002	.009	-.002	.011	.000	.013	-.000	.007	.002	.011
7-10	Total	.039	.049	.042	.054	.041	.051	.044	.057	.037	.046	.038	.050
	Talent	.039	.042	.039	.043	.043	.040	.044	.042	.036	.038	.036	.038
	Exper.	.000	.007	.003	.012	-.002	.011	.000	.015	.001	.009	.002	.012
11+	Total	.042	.053	.045	.064	—	—	—	—	.039	.052	.038	.045
	Talent	.039	.044	.039	.048	—	—	—	—	.037	.039	.037	.029
	Exper.	.003	.009	.005	.016	—	—	—	—	.003	.013	.001	.017

Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled “Actual”) or a randomly selected feasible allocation (columns labeled “Random”) among all school-year-field combinations with the number of eligible teachers specified by the row label in the field specified by the column label. The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the components of the gains attributable to task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t - 1$ prior to simulated reassignment in year t . *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t - 1$. See Section 3.7.1 and Appendix Section F for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data. Estimates of gains from task-specific experience and of teachers’ task-specific talent are derived from the Full Specification (equation (3.13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher’s task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.

Figure F.1: Tests for Dynamic Course Assignment Responses to Unobserved Time-Varying Endogenous Inputs



Notes: Figures display average school-teacher-year residuals (Figures F.1a and F.1e), school-teacher-subject-year residuals (Figures F.1b and F.1f), school-teacher-level-year residuals (Figures F.1c and F.1g), and school-teacher-subject-level-year residuals (Figures F.1d and F.1h), respectively, in the years leading up to a change in classroom assignment (using residuals from the *Restricted Specification*). *Restricted Specification* refers to a specification in which the school-teacher-subject-level fixed effects μ_{srl} from Equation (3.2) are restricted to be common across subject-levels within a school-teacher combination: $\mu_{srl} = \bar{\mu}_{st} \forall (j, l)$ and (s, r) . A permanent change in general course assignment (F.1e) is defined as a teacher-year combination in which the teacher is not observed teaching any course in a subsequent sample year. A permanent change in subject assignment (F.1f) is defined as a teacher-subject-year combination in which the teacher teaches the chosen subject, but is not observed teaching the chosen subject again in subsequent sample years. Permanent changes in level (F.1g) and subject-level (F.1h) assignments are defined analogously to permanent subject changes. Temporary changes in assignment are defined in a similar manner as permanent course assignment changes, except the teacher is observed returning to teach (F.1a) or teach in the chosen subject (F.1b), level (F.1c), or subject-level (F.1d) in a subsequent sample year. Bootstrap standard errors (in brackets) are computed using 1,000 iterations.

Table F.22: Counterfactual Simulations: Fraction of Classrooms Reallocated (Year-Based Measure of Experience, Excluding Teachers Without Full Histories)

Eligible Teachers	Math		Science		Social Studies	
	Static (1)	Dynamic (2)	Static (3)	Dynamic (4)	Static (5)	Dynamic (6)
2	0.253	0.303	0.230	0.329	0.267	0.317
3	0.331	0.391	0.337	0.412	0.370	0.444
4	0.404	0.449	0.410	0.482	0.422	0.479
5-6	0.428	0.469	0.434	0.505	0.463	0.518
7-10	0.427	0.469	0.484	0.574	0.488	0.542
11+	0.456	0.519	0.000	0.000	0.500	0.540

Notes: Each cell presents the fraction of classroom assignments in which a reallocation takes place (i.e. the simulated teacher assignment does not match the actual teacher assignment) among all school-year-field combinations with the number of eligible teachers specified by the row label in the field specified by the column label. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t - 1$ prior to simulated reassignment in year t . *Dynamic* refers simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t - 1$. See Section 3.7.1 and Appendix Section F for further detail about simulation methodology. A teacher is eligible for reassignment if their full teaching history is observed in the data.

Table F.23: Counterfactual Simulations: Achievement Gains from Optimal Allocation Relative to Actual and Random Allocations (Year-Based Measure of Experience, Including Teachers Without Full Histories)

	Eligible Teach.	Static		Dynamic	
		Actual (1)	Random (2)	Actual (3)	Random (4)
2	Total	.005	.010	.005	.011
	Talent	.004	.005	.004	.005
	Exper.	.001	.005	.000	.006
3	Total	.011	.018	.011	.021
	Talent	.010	.010	.011	.011
	Exper.	.001	.008	.001	.010
4	Total	.014	.023	.014	.027
	Talent	.012	.014	.012	.014
	Exper.	.001	.010	.001	.013
5-6	Total	.018	.030	.019	.034
	Talent	.017	.018	.017	.018
	Exper.	.001	.011	.002	.016
7-10	Total	.022	.034	.023	.040
	Talent	.020	.021	.020	.021
	Exper.	.002	.013	.003	.019
11+	Total	.023	.035	.025	.042
	Talent	.020	.021	.020	.021
	Exper.	.003	.014	.004	.021

Notes: Each cell presents simulated achievement gains from the optimal allocation of teachers to classrooms relative to either the observed allocation (in columns labeled “Actual”) or a randomly selected feasible allocation (columns labeled “Random”) among all school-year-field combinations with the number of eligible teachers specified by the row label. Classroom-level gains are pooled across the three fields (math, science, and social studies). The top entry in each cell displays the total achievement gains, while the middle and bottom entries display the contribution to this total of gains from task-specific experience and task-specific talent, respectively. *Static* refers to simulations in which teacher experience stocks are held fixed as they were in the actual sample through year $t - 1$ prior to simulated reassignment in year t . *Dynamic* refers to simulations in which teacher experience stocks used as the basis for simulated reassignment in year t are based on simulated assignments from 1995 through year $t - 1$. See Section 3.7.1 and Appendix Section F for further detail about simulation methodology. Eligible teachers consist of teachers who taught a test subject in the chosen school-year-field. Teachers who begin teaching prior to 1995 for whom full teaching histories were not observed are assigned imputed teaching histories as of 1995. See Section F for a description of the imputation procedure. Estimates of gains from task-specific experience and of teachers’ task-specific talent are derived from the Full Specification (presented in equation (3.13)). The principal incorporates information from empirical Bayes posterior beliefs about each teacher’s task-specific talent based on our school-teacher-subject-level fixed effect estimates for any school-teacher-subject-level combination that is observed in our sample. We assign task-specific productivities of 0 for any school-teacher-subject-level combination that we do not observe.